

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:05:44 ; Search time 1.09218 Seconds
(without alignments)
1830.070 Million cell updates/sec

Title: US-09-806-703A-12

Perfect score: 74

Sequence: 1 QYIKANSKFIGITEL 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A Geneseq 101002.*

1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	74	100.0	15	11	AA06310
2	74	100.0	15	18	AAW35506
3	74	100.0	15	18	AAW11505
4	74	100.0	15	19	AAW67033
5	74	100.0	15	19	AAW71321
6	74	100.0	15	20	AAW04051
7	74	100.0	15	20	AAW65758
8	74	100.0	15	20	AAW73220
9	74	100.0	15	21	AAW45511
10	74	100.0	15	21	AAW82637
					Tetanus toxin epit
					Universal T-cell e
					Tetanus toxoid uni-
					Universal helper T
					T-Helper epitope f
					T-cell epitope pep
					Tetanus toxoid epi
					Tetanus P2 epitope
					Tetanus toxoid T c

11	74	100.0	15	21	AAW92625	Foreign epitope P2
12	74	100.0	15	21	AAW84427	Amino acid sequenc
13	74	100.0	15	21	AAW70300	Clostridium tetani
14	74	100.0	15	21	AAW44763	Tetanus toxoid pro
15	74	100.0	15	22	AAE11763	Clostridium tetani
16	74	100.0	15	22	AAW99515	Vaccine related MH
17	74	100.0	15	22	AAW85701	Amino acid sequenc
18	74	100.0	15	22	AAW85451	Wild-type T830 (t
19	74	100.0	15	22	AAW61956	Tetanus Toxoid uni
20	74	100.0	15	22	AAW20143	Tetanus toxin T-ce
21	74	100.0	15	22	AAW68636	HER-2 B cell pepti
22	74	100.0	15	22	AAW46172	Tetanus toxoid T8
23	74	100.0	15	22	AAW49071	Tetanus toxoid TT
24	74	100.0	15	23	AAW97872	Tetanus toxin P2 (
25	74	100.0	16	18	AAW35445	T-cell stimulatory
26	74	100.0	16	20	AAW29705	Clostridium tetani
27	74	100.0	16	23	AAW93865	Clostridium tetani
28	74	100.0	16	23	AAW11413	Tetanus toxoid pre
29	74	100.0	17	15	AAW62692	Helper T cell epit
30	74	100.0	17	16	AAW82573	Tetanus toxin help
31	74	100.0	17	17	AAW05599	Tetanus toxin help
32	74	100.0	17	17	AAW88395	T-cell antigen TT2
33	74	100.0	17	21	AAW99274	HLA class II bindi
34	74	100.0	17	21	AAW80856	Pathogen derived T
35	74	100.0	17	21	AAW54539	T helper cell (Th)
36	74	100.0	17	21	AAW58768	Unidentified pepti
37	74	100.0	17	22	AAW99516	Vaccine related MH
38	74	100.0	17	22	AAW62904	Amino acid residu
39	74	100.0	17	22	AAW84435	Amino acid sequenc
40	74	100.0	17	22	AAW30941	Amino acid sequenc
41	74	100.0	17	22	AAW31029	Antigenic fragment
42	74	100.0	17	22	AAW31118	Antigenic fragment
43	74	100.0	17	22	AAW15589	Peptide 5 for pept
44	74	100.0	18	20	AAW26607	HIV-derived lipope
45	74	100.0	18	23	AAW809794	Peptide TT functio

ALIGNMENTS

RESULT 1
AAW06310
ID AAR06310 standard; protein; 15 AA.
XX
AC AAR06310;
XX
04-DSC-1990 (first entry)
XX
DE Tetanus toxin epitope.
XX
KW Tetanus toxin; vaccine; major histocompatibility complex; MHC;
KW antimalarial.
XX
OS Synthetic.
XX
PN EP378881-A.
XX
PD 25-JUL-1990.
XX
PF 27-DEC-1989; 89EP-0203318.
XX
PR 16-NOV-1989; 89IT-0022409.
PR 17-JAN-1989; 89IT-0019110.
XX
(ENIE) ENRICHES SPA.
PA Pessi A, Bianchi E, Verdini AS, Corradin G;
PI
DR WPI; 1990-225582/30.
XX
PT Synthetic peptide(s) corresp. to tetanus toxin epitope(s) - used
PT as universal carriers for prepn. of immunogenic conjugate(s) for
PT use as vaccines.

THIS PAGE BLANK (USPTO)

XX
PS
CC
CC
CC
CC
CC
CC
CC
CC
SQ

Claim 1; Page 17; 20pp; English.

Epitopic peptides may be used with synthetic haptens derived from a pathogen to generate an immune response to the pathogen. Peptides are recognised by numerous T-helper cell clones within the context of a wide range of alleles of the human MHC. The peptides may be used in an antimalarial vaccine inducing Ab. response to P.falciparum.

SQ Sequence 15 AA;

Query Match 100.0%; Score 74; DB 11; Length 15;
Best Local Similarity 100.0%; Pred. No. 3.9e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 QYIKANSKFIGITEL 15
DB 1 QYIKANSKFIGITEL 15

RESULT 2

AAW35506
ID AAW35506 standard; peptide; 15 AA.

XX AC AAW35506;

XX DT 22-APR-1998 (first entry)

XX DE Universal T-cell epitope peptide SEQ ID NO:8.

XX KW T-cell stimulatory peptide; immunogen; non-dendritic; carrier; tumour;
XX KW scaffold; inhibition; metastasis; wound healing; solid phase.

XX OS Unidentified.

XX PN WO9738011-A1.

XX PD 16-OCT-1997.

XX PF 03-APR-1997; 97WO-DE00146.

XX PR 03-APR-1996; 96DK-0000398.

XX PP (PEPR-) PEPRESEARCH AS.

XX PI Heegaard PMH, Jakobsen PH;

XX DR WPI; 1997-512645/47.

XX PT Non-dendritic peptide carrier linked to a solid phase - useful as a
XX PT diagnostic agent and as a scaffold for production of chemical
XX PT derivatives

XX PS Example 20; Page 124; 262pp; English.

XX CC A non-dendritic peptide carrier (A) has been developed which is coupled
XX CC through a linker to a solid phase, forming a complex of (A)-solid phase.
XX CC Where (A) comprises 10-50 amino acids capable of forming a secondary
XX CC structure in a benign buffer after liberation from the solid phase, and
XX CC further the (A)-solid phase complex comprises an immunogenic substance
XX CC and/or an immune mediator coupled on (A). The present sequence
XX CC represents a peptide used in an example from the present invention. An
XX CC (A)-solid phase complex can be used as a scaffold for the production of
XX CC chemical derivatives, characterised by covalently attaching molecules at
XX CC these attachment points. Alternatively (A) is used as a scaffold-peptide for
XX CC incorporation into an immunostimulating Complex (Iscom) resulting in
XX CC an Iscom complex which is used for the chemical coupling of antigenic
XX CC peptides having fibronectin-, laminin- or vitronectin-like
XX CC structures. The complex is used for the promotion of cell-attachment to
XX CC particular to inhibit tumour growth and metastasis,
XX CC wound healing. Also a derivatised (A) can be used

CC
CC
CC
CC
CC
SQ

for the selection of specifically-binding aptamers or as a diagnostic agent. Such diagnostic-(A) molecules could be used to detect molecules derived from or indicative of pregnancy or of a disease, such as an infectious, autoimmune or cancerous disease.

SQ Sequence 15 AA;

Query Match 100.0%; Score 74; DB 18; Length 15;
Best Local Similarity 100.0%; Pred. No. 3.9e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 QYIKANSKFIGITEL 15
DB 1 QYIKANSKFIGITEL 15

RESULT 3

AAW11505
ID AAW11505 standard; Protein; 15 AA.

XX AC AAW11505;

XX DT 24-SEP-1997 (first entry)

XX DE Tetanus toxoid universal Th epitope TT830.

XX KW Humanised antibody; anti-Fc receptor; H22; bifunctional; bispecific;
XX KW fusion protein; chimera; tetanus toxoid; helper T cell epitope;
XX KW antigen presentation; ds.

XX OS Clostridium tetani.

XX PN WO9640789-A1.

XX PD 19-DEC-1996.

XX PF 07-JUN-1996; 96WO-US09988.

XX PR 07-JUN-1995; 95US-0484172.

XX PP (MEDA-) MEDAREX INC.

XX PI Deo YM, Goldstein J, Graziano R, Somasundaram C;

XX DR WPI; 1997-052242/05.

XX DR N-PSDB; AAT58127.

XX PT Recombinant, multi-specific anti-Fc receptor antibody molecules -
XX PT also comprise an anti-target portion, used for the treatment of
XX PT cancer, autoimmune disease and pathogenic infection

XX PS Example 7; Fig 24; 115pp; English.

XX CC Synthetic DNA coding for the wild-type universal Th epitope from
XX CC tetanus toxoid, designated TT830, was fused to the 3'-end of DNA
XX CC encoding heavy chain sequences from the humanised anti-Fc gamma RI
XX CC monoclonal antibody H22. The resulting fusion protein was shown to
XX CC be significantly more efficient in antigen presentation and T cell
XX CC stimulation than the TT830 epitope alone. A similar fusion
XX CC construct was prepared coding for a mutant, antagonistic form of the
XX CC epitope (designated TT833S) fused to the anti-Fc gamma RI. The
XX CC Fab22-TT833S is at least 100 times more effective than TT833S in
XX CC inhibiting T cell activation.

SQ Sequence 15 AA;

Query Match 100.0%; Score 74; DB 18; Length 15;
Best Local Similarity 100.0%; Pred. No. 3.9e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 QYIKANSKFIGITEL 15

DB 1 QYIKANSKFIGITEL 15

THIS PAGE BLANK (USPTO)

us-09-806-703a-12.rag

012 2003

26-NOV-1998 (first entry)
 Universal helper T-cell epitope P2 derived from tetanus toxin.
 Liver stage; Plasmodium; Navy Yoelii Liver Stage 3 antibody; NYLS3;
 hepatic and erythrocytic stage protein; PyHEP17; vaccine;
 malaria parasite; tetanus toxin; P2; helper T-cell epitope.
 Synthetic.
 Clostridium tetani.
 US5814617-A.
 29-SEP-1998.
 07-OCT-1994; 94US-0319704.
 07-OCT-1994; 94US-0319704.
 (USNA) US SEC OF NAVY.
 Charoenvit Y, Doolan DL, Hedstrom RC, Hoffman SL;
 WPI; 1998-541794/46.
 Vaccine for protecting mammal against infection by malaria caused by
 Plasmodium species - comprises a first nucleic acid encoding a first
 polypeptide capable of eliciting an immune reaction against an
 antigen expressed during the liver
 Disclosure; Column 12; 24pp; English.
 AAW71321-22 represent universal helper T-cell epitopes derived from
 tetanus toxin. They are used to enhance host immune response to
 vaccines. The specification describes a Plasmodium yoelii liver stage
 17 kDa hepatic and erythrocytic stage protein designated PyHEP17. This
 protein elicits a response from an Igi monoclonal antibody designated
 CC Navy Yoelii Liver Stage 3 (NYLS3). This antibody does not recognise
 sporozoites, but does recognise P. yoelii liver stage parasites. NYLS3
 eliminates upto 90% of liver stage parasites. The specification
 describes a vaccine for reducing the severity or incidence of infection
 by a malaria parasite of the genus Plasmodium. The DNA vaccine comprises
 exon 1 and part of exon 2 of the PyHEP17 gene.
 Query Match 100.0%; Score 74; DB 19; Length 15;
 Best Local Similarity 100.0%; Pred. No. 3.9e-07;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 QYIKANSKFIGITEL 15
 DB 1 QYIKANSKFIGITEL 15
 RESULT 6
 AAW04051
 ID AAW04051 standard; peptide; 15 AA.
 XX AC AAW04051;
 XX DT 04-JAN-2000 (first entry)
 XX DE T-Helper epitope from tetanus toxoid.
 XX KW Covalently reactive antigen analog; CRAA; catalytic antibody;
 KW electrophilic reaction centre; phosphonate; boronate; vaccine;
 KW transition state analog; TSA; isostere; gp120; HIV-1; T-helper;
 KW tetanus; toxoid; B-T-epitope.
 XX OS Clostridium tetani.
 XX PN WO9948925-A1.

DT XX 26-NOV-1998 (first entry)
 DE XX Universal helper T-cell epitope P2 derived from tetanus toxin.
 XX KW Liver stage; Plasmodium; Navy Yoelii Liver Stage 3 antibody; NYLS3;
 KW hepatic and erythrocytic stage protein; PyHEP17; vaccine;
 KW malaria parasite; tetanus toxin; P2; helper T-cell epitope.
 XX OS Synthetic.
 XX OS Clostridium tetani.
 XX PN US5814617-A.
 XX PD 29-SEP-1998.
 XX PF 07-OCT-1994; 94US-0319704.
 XX PR 07-OCT-1994; 94US-0319704.
 XX PA (USNA) US SEC OF NAVY.
 XX PI Charoenvit Y, Doolan DL, Hedstrom RC, Hoffman SL;
 XX DR WPI; 1998-541794/46.
 XX PT Vaccine for protecting mammal against infection by malaria caused by
 PT Plasmodium species - comprises a first nucleic acid encoding a first
 PT polypeptide capable of eliciting an immune reaction against an
 PT antigen expressed during the liver
 XX PS Disclosure; Column 12; 24pp; English.
 XX CC AAW71321-22 represent universal helper T-cell epitopes derived from
 CC tetanus toxin. They are used to enhance host immune response to
 CC vaccines. The specification describes a Plasmodium yoelii liver stage
 CC 17 kDa hepatic and erythrocytic stage protein designated PyHEP17. This
 CC protein elicits a response from an Igi monoclonal antibody designated
 CC CC Navy Yoelii Liver Stage 3 (NYLS3). This antibody does not recognise
 CC sporozoites, but does recognise P. yoelii liver stage parasites. NYLS3
 CC eliminates upto 90% of liver stage parasites. The specification
 CC describes a vaccine for reducing the severity or incidence of infection
 CC by a malaria parasite of the genus Plasmodium. The DNA vaccine comprises
 CC exon 1 and part of exon 2 of the PyHEP17 gene.
 SQ Sequence 15 AA;
 Query Match 100.0%; Score 74; DB 19; Length 15;
 Best Local Similarity 100.0%; Pred. No. 3.9e-07;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 QYIKANSKFIGITEL 15
 DB 1 QYIKANSKFIGITEL 15
 RESULT 5
 AAW71321
 ID AAW71321 standard; peptide; 15 AA.
 XX AC AAW71321;
 XX XX

THIS PAGE BLANK (USPTO)

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	112	100.0	21	12	AAR11896	Immunogenic conjug
2	112	100.0	21	17	AAW06130	Tetanus toxoid pro
3	112	100.0	21	17	AAR088397	T-cell antigen TT3
4	112	100.0	21	18	AAW46449	Broad range helper
5	112	100.0	21	19	AAW67034	Tetanus toxin frag
6	112	100.0	21	20	AAW67579	Tetanus toxin frag
7	112	100.0	21	20	AAW73222	T-cell epitope pep
8	112	100.0	21	21	AAW73222	Tetanus toxoid epi
9	112	100.0	21	21	AAW45512	Tetanus P30 epitop
10	112	100.0	21	21	AAW99876	Tetanus toxin T ce
					AAW92636	Foreign epitope P3

Synthetic peptide(s) used as universal carriers - for preparing

THIS PAGE BLANK (USPTO)

Tue Jul 22 12:43:19 2003

us-09-806-703a-14.rag

PT immunogenic conjugates used as vaccines against Plasmodium
PT falciparum
XX
PS Claim 1; page 13; 16pp; English.
XX
CC This peptide corresps. to residues 947-967 of Tetanus toxin. It can
CC be used as a universal carrier for the prepn. of an immunogenic
CC conjugate. It is covalently bound to a peptide or polysaccharide
CC hapten derived from a pathogen. This conjugate can be used as
CC a vaccine for malaria. This peptide is recognised by different T-
CC helper cell clones in association with alleles of the human MHC.
CC It contains 2 epitopes: (a) 953-967, recognised by DR5-restricted
CC clones; and (b) 947-960, recognised by all other DR and DP-
CC restricted clones.
XX
SQ Sequence 21 AA;

Query Match 100.0%; Score 112; DB 12; Length 21;
Best Local Similarity 100.0%; Pred. No. 3.9e-12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FNNFTVSFWLRVPKVSASHLE 21
Db 1 FNNFTVSFWLRVPKVSASHLE 21

RESULT 2
AAW06130
ID AAW06130 standard; Peptide; 21 AA.
XX
AC AAW06130;
XX
DT 07-FEB-1997 (first entry)
XX
DE Tetanus toxoid protein T-cell epitope.
XX
KW Cholesteryl ester transfer protein; CETP; antigen; vaccine;
KW cardiovascular disease; atherosclerosis; tetanus toxoid;
KW T-cell epitope.
XX
OS Clostridium tetani.
XX
PN WO9634888-A1.
XX
PD 07-NOV-1996.
XX
PF 01-MAY-1996; 96WO-US06147.
XX
PR 01-MAY-1995; 95US-0432483.
XX
PA (TCEL-) T CELL SCI INC.
XX
PI Rittershaus CW, Thomas LJ;
XX
DR WPI; 1996-506103/50.
XX
PT Cholesteryl ester transfer protein B cell epitope linked to T cell
PT epitope - used to generate vaccine to regulate CETP activity for
PT decreasing the risk of developing a cardiovascular disease e.g.
PT atherosclerosis
XX
PS Claim 11; Page 43; 72pp; English.

XX
CC A helper T-cell epitope (AAW06130) comprises amino acids 947-967
CC of tetanus toxoid protein. It can be utilised in novel peptide
CC vaccines (see also AAW06129, AAW06132) also including B-cell
CC epitopes from human or rabbit cholesteryl ester transfer
CC protein. The peptide is used to generate an immune response against endogenous
CC cholesteryl ester transfer protein or preventing a cardiovascular
CC disease.

Query Match 100.0%; Score 112;
Best Local Similarity 100.0%; Pred. No.
Matches 21; Conservative 0; Mismatch

Qy 1 FNNFTVSFWLRVPKVSASHLE 21
Db 1 FNNFTVSFWLRVPKVSASHLE 21

RESULT 3
AAR88397
ID AAR88397 standard; Peptide; 21 AA.
XX
AC AAR88397;
XX
DT 12-JUN-1996 (first entry)
XX
DE T-cell antigen TT3 peptide.
XX
KW T-antigen; vaccine; antibody; T-cell; T-
KW alpha-helix; coiled-coil heterodimer; cc
XX
OS Synthetic.
XX
PN WO9531480-A1.
XX
PD 23-NOV-1995.
XX
PF 18-MAY-1995; 95WO-CA00293.
XX
PR 18-MAY-1994; 94US-0245507.
XX
PA (SPIS-) SPI SYNTHETIC PEPTIDES INC.
XX
PI Cachia PJ, Hodges RS, Houston ME, Irv
PI Zhou NE;
XX
DR WPI; 1996-010880/01.
XX
PT Hetero:dimeric polypeptide immunogen in
PT with different antigens on each sub:unit
PT for antibody prodn.
XX
PS Claim 7; Page 62; 95pp; English.
XX
CC This T-cell antigen TT3 peptide may be
CC contained in one of the 2 subunits of ar
CC heterodimer. Each core peptide is comp
CC internal AA repeat sequences. This pept
CC to the core peptide through covalent lin
CC internal repeats. The 2 subunits of the
CC in a stable alpha-helical coiled-coil co
CC stoichiometry, and the peptide antigen i
CC surfaces of the configuration. The hete
CC synthetic vaccine (optionally multivaler
CC antibodies.
XX
SQ Sequence 21 AA;

Query Match 100.0%; Score 112;
Best Local Similarity 100.0%; Pred. No.
Matches 21; Conservative 0; Mismatch

Qy 1 FNNFTVSFWLRVPKVSASHLE 21
Db 1 FNNFTVSFWLRVPKVSASHLE 21

RESULT 4
AAW46449
ID AAW46449 standard; Peptide; 21 AA.
XX
AC AAW46449;

THIS PAGE BLANK (USPTO)

Db 1190 SASSG-----PPKAEEDYVNEPLNTFTNALGNAEYMKNSLLSVPEKAKAFDNDPDY 1243
 Qy 1230 QDDPPERGA--PPSTFKGTPT-----AENPEYL 1255
 Db 1244 NLSLPPRSTLQHPDYLQBYSTKYFYKQNGRIRPIVAENPEYL 1285

RESULT 8
 XMRK XIPMA
 ID XMRK XIPMA STANDARD; PRT; 1167 AA.
 AC P13388;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 16-OCT-2001 (Rel. 40, last sequence update)
 DT 15-JUN-2002 (Rel. 41, last annotation update)
 DE Melanoma receptor protein-tyrosine kinase precursor (EC 2.7.1.112).
 GN XMRK OR TU.
 OS Xiphophorus maculatus (Southern platyfish).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
 OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
 OX NCBI_TaxID=8083;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90015140; PubMed=2797166;
 RA Wittbrodt J., Adam D., Malitschek B., Maueier W., Raulf F.,
 RA Telling A., Robertson S.M., Scharltl M.;
 RA "Novel putative receptor tyrosine kinase encoded by the melanoma-
 RA inducing Tu locus in Xiphophorus.";
 RL Nature 341:415-421(1989).
 RN [2]
 RP REVISION TO 515.
 RA Scharltl M.;
 RA Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 RL -!- FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.
 CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.
 CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X16891; CAA34770.2; -;
 CC PIR; S06142; S06142.
 CC HSSP; P11362; 1FGK.
 CC InterPro; IPR000494; EGF_L domain.
 CC InterPro; IPR000719; Euk_pkinase.
 CC InterPro; IPR002174; Furin-like.
 CC InterPro; IPR002290; Ser_thr_pkinase.
 CC InterPro; IPR01245; Tyr_pkinase.
 CC Pfam; PF00069; pkinase; 1.
 CC Pfam; PF00757; Furin-like; 1.
 CC Pfam; PF01030; Recep_L domain; 2.
 CC PRINTS; PR00109; TYRKINASE.
 CC ProDom; PD000001; Euk_pkinase; 1.
 CC SMART; SM00261; FU; 5
 CC SMART; SM00220; S_TKc; 1.
 CC SMART; SM00219; Tyrc; 1.
 CC PROSITE; PS00107; PROTEIN KINASE ATP; 1.
 CC PROSITE; PS00109; PROTEIN KINASE TYR; 1.
 CC PROSITE; PS50011; PROTEIN KINASE_DOM; 1.
 CC Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
 KW Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.
 FT SIGNAL 1 25
 FT CHAIN 26 1167 MELANOMA RECEPTOR PROTEIN-TYROSINE

FT DOMAIN 26 642
 FT TRANSMEM 643 665
 FT DOMAIN 666 1167
 FT DOMAIN 710 977
 FT NP_BIND 716 724
 FT BINDING 743 743
 FT ACT_SITE 835 335
 FT DISULFID 195 204
 FT DISULFID 199 212
 FT DISULFID 220 228
 FT DISULFID 224 236
 FT DISULFID 237 245
 FT DISULFID 241 253
 FT DISULFID 256 265
 FT DISULFID 269 296
 FT DISULFID 300 311
 FT DISULFID 315 330
 FT DISULFID 333 337
 FT DISULFID 504 513
 FT DISULFID 508 521
 FT DISULFID 524 533
 FT DISULFID 537 553
 FT DISULFID 556 569
 FT DISULFID 560 577
 FT DISULFID 593 615
 FT DISULFID 618 626
 FT DISULFID 622 634
 FT CARBOHYD 114 114
 FT CARBOHYD 144 144
 FT CARBOHYD 201 201
 FT CARBOHYD 356 356
 FT CARBOHYD 365 365
 FT CARBOHYD 398 398
 FT CARBOHYD 417 417
 FT CARBOHYD 501 501
 FT CARBOHYD 576 576
 FT CARBOHYD 621 621
 SQ SEQUENCE 1167 AA; 129934 MW; 4793E0749DC1D55A CRC64;

Query Match 38.7%; Score 2649.5; DB 1; Length 1167;
 Best Local Similarity 44.6%; Pred. No. 6.7e-134;
 Matches 570; Conservative 164; Mismatches 390; Indels 155; Gaps 30;

Qy 4 AALRCWGLLALLPPGAASL---QVCTGTDMKLRIPASPEHLDMLRHLYQGVVQGN 59
 Db 8 AALLQ--LLLVLSIRCCSTDPBRKVCGTSTNQMTM---LDNHYLKMVKMYSGCNVLEN 62
 Qy 60 LEITYPTNASLSFLQDIQEV-----QFNFTVSFWLRVPKVSASHLEORLRIVRGTL 113
 Db 63 LEITYTQENQDLSFLQSIQEVGGVLIAMNEVST-----IPLVN-----LRLIRGNL 110
 Qy 114 FEDNYALAVLDNGDPLNNTTPTVGASPGGLRELQLSLTEILKGVLIQRNPOLCYODTI 173
 Db 111 YEGNFTLLVMSNYQK-PPSSP--DVYQVGLKQLQSLNLTLEILSGGVKVSHPNLLCNVETI 167
 Qy 174 LKWDIFHKNNQALTLTIDNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVCAAGC-AR 232
 Db 168 NWDIVDKTSNPTNMLIPHAFERQCCKDHGCVNGSCWAPGPHGCKFTLLCAEQCNRR 227
 Qy 233 CKGPLPTDCCCHQCAAGCTGPKHSDCILACLFHNSGICELHCPALVTYNTDTFESMPNPE 292
 Db 228 CRGPKPIDCCNEHCAGCTGPRATDCLACRDFNDGDKCTCPPPKIYDIVSHQVVDNPN 287
 Qy 293 GRITFGASCVTACPNYVLTSDVGSCTLVCPHMQEVTAEADGTORCEKCSKPCARVCVGLG 352
 Db 288 IKYTFGAACVKCEPSNVVTE-GACVRSKCSAGMLEVD-ENGKRSCKPCDGVCPKVCDDIG 345
 Qy 353 MEHLREVRAVTSANIOEFACKKIFGSLAFLPESFEDGRASNTAPLOPEOLQVPELSEI 412
 Db 346 IGSLSNTIAVNSNIRSFNSCTKINGDIILNRNSFEGDPHYKIGTMDPEHLNLTUTTKEI 405
 Qy 413 TGYLYISAWPDSLPDLSVFQNLQVIRGRIHLNGAYS--LTQLGLGISWLRLSRELGSGL 471

```
Db 406 TGVLVIMWMPENMTSLVFQNLRIIRGRTTSRGFSFVVQVRHLQWLGLSLAEVSAGN 465
Qy 472 ALIHNTLFCVHTVVDQDLFRNPHQALLHTANRPEDCEVGEGLACHOLCARGHCWPGP 531
Db 466 VILKNTLQLRANTINMRRLFRSEDSQSTIEYDART-----ENQTCNNECEDGCGWGP 518
Qy 532 TQCYNCSQFLRGQCEVCRVQLGLPREYVVARHCLPCHPCQCPQNGSVTCFGEADQCV 591
Db 519 TMCVSLHVDGRGRCVASCNLQGEPREAQVDGRCVQCHQCLVQTSLTCTCYGPGPACS 578
Qy 592 ACAHYKDPFFCVARCFSGKPDLSYMTIWKPFDEEGACQPCFINCTHSCVDLDDKGCFAE 651
Db 579 KSAHFQDGPQICPRCPHGILGDGTL-LWKYADRWGQCQPCQHCNCTQCSGPGSLGCRGD 637
Qy 652 QRAEPLTSIVSAVGVILLVVLGVVFGILLIKRQOKIRKYMRRLLOBTVELVBLTSGA 711
Db 638 -IVSHSLAVGLVSGLLITVIVALLVVLVLRRIK-RKRTIRCLQEKELVEPLTSGQ 695
Qy 712 MPNOAQRILKETELRKVKVLGSGAGFYVYKGIWIPDGENVKIPVAIKVLRENTSPKANK 771
Db 696 APNQAFLIKETEFKORVLGSGAGFYVYKGIWIPDGENIRIPVAIKVLRENTSPKANK 755
Qy 772 EILDEAYVMAGVSPYSRLGICLTSTVQLVTQMLPYGCLLDVHRENRGLSGQDLNW 831
Db 756 EVLDEAYVMSVDHPVHCRLGICLTSAVQLVTQMLPYGCLLDVYRQHCIRICGQWLLNW 815
Qy 832 CMQIAKMSYLEDVRLVHRDLAARNLVKSNHVKITDFGLARLLDDDETHYHADGGKVP 891
Db 816 CVQIAKGMNLEERHLVHRDLAARNLVKSNHVKITDFGLSKLITADEKEYQADGGKVP 875
Qy 892 IKWMALESILRRRTHQSDVMSYGVYVWELMTFGAKPYDGPAREIPDLLEKGERLPPOP 951
Db 876 IKWMALESILQWYTHQSDVMSYGVYVWELMTFGSKPYDGPAREIASVLENGERLPPOP 935
Qy 952 ICTIDVIMVYKMWIDSECRPRELVSERFMRARDPQREVVIONEDLGPASPLDSTFY 1011
Db 936 ICTIEVYMIILKMWIDSPSRPRELVSERFMRARDPQREVVIONEDLGPASPLDSTFY 992
Qy 1012 RSLLEDDMGDLVDAEYLVQQGFCDPAPAGAGGWVHRRSSSTRSGGDLTLGLPEP 1071
Db 993 SRLLSDD--DVVDAEYLLPYKRI-----NRQGS----- 1020
Qy 1072 SEEAAPSLAPSEGAGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTV-PLPSET 1130
Db 1021 -----EPCIPPTGH-----PVRENSITLRNISDFTQNALEKDL 1053
Qy 1131 DGYVAPLTCSPQPEYVNPQDVRPQ-----PSPRE-----GPIP-AARPAGATLERAK 1177
Db 1054 DGH-----EYVNPQGETSSRLSDIYNPNVEDLTDGMPVSLSSQEAETNFSRPE 1103
Qy 1178 TISPCKNGVVDVFAFGGAVENPEYLTQGGAAQPPHPPAFSPAFDNLVWDQDPBPG 1237
Db 1104 YLNTNQSL-----PLVSSGSMDDPY-----YQAAF-----LPQTG 1138
Qy 1238 APPSTFKTPTAENPEYLG 1256
Db 1139 ALTGNGMFLPAENLEYLG 1157
```

RESULT 9

ERB3_HUMAN

ID ERB3_HUMAN STANDARD; PRT; 1342 AA.

AC P21860;

DT 01-MAY-1991 (Rel. 18, Created)

DT 01-MAY-1991 (Rel. 18, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)

DE (c-erbB3) (Tyrosine kinase-type cell surface receptor HER3).

CN ERB3 OR HER3

OS Homo sapiens (Human)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NCBI_TaxID=9606;

[1]

SEQUENCE FROM N.A.

MEDLINE=90093234; PubMed=2687875;

Kraus M.H., Issing W., Miki T., Popescu N.C., Aaronson S.A.;

"Isolation and characterization of ERBB3, a third member of the

ERBB/epidermal growth factor receptor family: evidence for

overexpression in a subset of human mammary tumors.";

Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).

[2]

SEQUENCE FROM N.A.

MEDLINE=90311312; PubMed=2164210;

Plowman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,

Todaro G.J., Shoyab M.;

"Molecular cloning and expression of an additional epidermal growth

factor receptor-related gene.";

Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909 (1990).

[3]

SEQUENCE FROM N.A. (SHORT FORM).

TISSUE=Placenta;

MEDLINE=93282822; PubMed=7685162;

Katoh M., Yazaki Y., Sugimura T., Terada M.;

"c-erbB3 gene encodes secreted as well as transmembrane receptor

tyrosine kinase.";

Biochem. Biophys. Res. Commun. 192:1189-1197(1993).

-1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTK.

-1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein

tyrosine phosphate.

-1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS

(POTENTIAL).

-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND

SECRETED (SHORT FORM).

-1- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A

SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.

-1- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.

-1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE

SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.

-1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES

AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF

PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).

-1- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.

-1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)

or send an email to license@isb-sib.ch.

EMBL; M29366; AAA35790.1; --

EMBL; M34309; AAA35979.1; --

EMBL; S61953; AAB26935.1; --

PIR; A36223; A36223.

HSSP; P11362; 1FGK.

Genew; HGNC:3431; ERBB3.

MIM; 190151; --

InterPro; IPR000494; EGFR_L_domain.

InterPro; IPR000719; Euk_pkinase.

InterPro; IPR002174; Furin-like.

InterPro; IPR001245; Tyr_pkinase.

Pfam; PF00069; pkinase; 1.

Pfam; PF00757; Furin-like; 1.

Pfam; PF01030; Recep_L_domain; 2.

ProDom; PD000001; Euk_pkinase; 1.

SMART; SMC0261; FU; 3.

SMART; SMC0219; TyKc; 1.

PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.

PROSITE; PS00109; PROTEIN_KINASE_TYR; FALSE_NEG.

PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.

Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;

Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;

[illegible]

```

RESULT 10
ERB3_RAT
ID ERB3_RAT STANDARD; PRT; 1339 AA.
AC Q62799; O62955;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (C-erbB3).
GN ERB3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBT_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=96096535; PubMed=8522190;
RA Hellyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Koland J.G.;
RT "Cloning of the rat ErbB3 cDNA and characterization of the
RT recombinant protein.";
RL Gene 165:279-284 (1995).
RN [2]
RP REVISIONS TO 85: 513 AND 565.
RA Hellyer N.J., Koland J.G.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 922-1097 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Frohnert P.W., Kim S.S., Corbett J.A.;
RT ErbB3, is induced during Wallerian degeneration.";
RL J. Neurosci. 17:1642-1659 (1997).
CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTA.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U29339; AAC28498.2; -
DR EMBL; U52530; AAC3050.1; -
DR HSSP; P11362; IFGK.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 5.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; FALSE_NEG.
DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
DR PROSITE; PS00011; PROTEIN KINASE DOM; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; signal;

```

```

KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 19
FT CHAIN 20 1339
FT DOMAIN 20 643
FT TRANSMEM 644 662
FT DOMAIN 663 1339
FT DOMAIN 183 259
FT DOMAIN 707 964
FT NP_BIND 713 721
FT BINDING 740 740
FT ACT_SITE 832 832
FT DISULFID 186 194
FT DISULFID 190 202
FT DISULFID 210 218
FT DISULFID 214 226
FT DISULFID 227 235
FT DISULFID 231 243
FT DISULFID 246 255
FT DISULFID 259 286
FT DISULFID 290 301
FT DISULFID 305 320
FT DISULFID 323 327
FT DISULFID 500 509
FT DISULFID 504 517
FT DISULFID 520 529
FT DISULFID 533 549
FT DISULFID 556 573
FT DISULFID 576 585
FT DISULFID 589 610
FT DISULFID 613 621
FT DISULFID 617 629
FT CARBOHYD 126 126
FT CARBOHYD 250 250
FT CARBOHYD 353 353
FT CARBOHYD 408 408
FT CARBOHYD 414 414
FT CARBOHYD 437 437
FT CARBOHYD 469 469
FT CARBOHYD 522 522
FT CARBOHYD 566 566
FT CARBOHYD 616 616
FT CONFLICT 1028 1028
SQ SEQUENCE 1339 AA; 147545 MW; 0AA5F2402BBDF1E CRC64;
Query Match 33.8%; Score 2316.5; DB 1; Length 1339;
Best Local Similarity 40.4%; Pred. No. 4.3e-116;
Matches 523; Conservative 170; Mismatches 429; Indels 173; Gaps 37;
Qy 3 LAALCRWGLLLALLPPGAA---STQVCTGTDMLRLPASPEHLDMLRLHYQGVQVQGN 59
Db 7 LQVLC-----FLLSLARGSEMGNSQAVCPGTLNGLSVTGADNNOYQTLKLYKEVWGN 62
Qy 60 LEITYLPTNASLSPLQDIQE-----VQFNFTVSFWLRVPKVSASHLEQRLRVRGTQL 113
Db 63 LEIVLTGHNADSLPQWIREVTGVVVMNEFSV---LPLP-----NLRVVRGTQV 110
Qy 114 FEDNYALAVLDNGDPINNTPTVTGASPGGLRELQRLSLTELLKGVLIQRPOLCYQDTI 173
Db 111 YDGKFAIFVM-----LNYNT-----NSSHALRQLKFTQTLTEILSGGVIERKNDKCHMDTI 161
Qy 174 LWKDIFHKNNQALTLDTNRSRACHPCSPMKSCRCWGESSEDCQSILRTVTCAGGC-AR 232
Db 162 DWRDIVRVR---GAEIVVKNKNGANCPCHVECKG-RWGGPGPDDCOLTKTICAPQCNCR 217
Qy 233 CKGELPTDCHEQCAAGCTGPKSHDCLACLHFNHSGICELHCPALVYNTDTTFSSMPNPE 292
Db 218 CFGENPNQCHDECAGCGSGPQDTCFACRFNDSGACVPRCPPLVYNKLTFLQLEPNPH 277
Qy 293 GRITFGASCVTACPNYLSLTDVSGCTLVCLPHNOEVAETAGDQTCRCKSKCARVCVCLG 352
Db 278 TKYQYGGVGVASCAPHNFV-VDQTFVCRACPPDKMEVD-KHGLKMKCEPCGGCLCPKACEGTG 335
Qy 353 MEHLREVRVTSANIQEFAGCKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFETLBEI 412

```

```
Db 336 SG- -SRYQTVSSNIDGVNCTKILGNLDFILITGLNVDPMHKIPALDPEKLNVRFTVREI 393
Qy 413 TGYLYISAWPDLPLDLSFQNLQVIRGRIHLNGAYS-LTLOGLGISWLGRLSLRELASGL 471
Db 394 TGYLNIQSWPHMNFVFSNLTIIGBSLYNRGFSLLIMKLNVTSLGFRSLKEISAGR 453
Qy 472 ALIHHTHLCFVHTVPMQDLPFRNHQALLHTA-NRPEDECVGEGGLACHQLCARGHCWGP 530
Db 454 VYISANQOLCYHHSNMTLLRLLRGPSEERLDIKIDRPLGECLEAGKVCDPGLSSGCGWGP 513
Qy 531 PTOCVNSQFLRGQECVEECRVLOGLPREYVNAHCLPCHPECOPNGSVTCFGEADQC 590
Db 514 PGQLCSKNYSREGVCTHCFNLOGEPRFVHEAQCFSCHPECLPMEGTSTCNGSGSDAC 573
Qy 591 VACAHYKDPFPCVARGSGVKPDLISYMPIWKFPEEGACQPCINCTHSC--VLDLDDKGC 648
Db 574 ARCAHFRDGHCVNSCPHGILG--AKGPIYKYPDAQNECRCHENCTGCGNGPELQDCLG 631
Qy 649 PAQRASPLTSIVSAVGIILVVLGVVFGVILIKRQOKIR-KYTMRLRLLOETELVPEPLT 707
Db 632 QAEVLSMKPHLVIAVTVG--LAVILMILGGSFLYWRGRIQNKAMRRYLERGESIEPLD 689
Qy 708 PSGAMPQAOAMRILKETELRKVKVLGSGAFQTVYKGIWIPGENVKIPVAIKVLENTSP 767
Db 690 PS-EKANKVLARIFKETELRKVLGSGVFTVHKGIWIPGESIKIPVCIKVIEDKSGR 748
Qy 768 KANKEILDEAYVMAGVSPYVSRLLGLCTSTVOLVTOLMPYVGLDHHVNRNRLGSGD 827
Db 749 QSFQAVTDHMLAVGSLDHAHIVRLGLGCPGSLQVLTQYLPGLSLLDHVKQHRETLGPQL 808
Qy 828 LLNMCQIAGMSYLEDLVRLVHRLDAARNLVKSPNHVKIDFGLARLLDIDETEHADG 887
Db 809 LLNMGVOIAKMYLEHSMVHRDLARNVLMKSPSOVQVADLVPPDDKQLLHSE 868
Qy 888 GKVPKMALESILRRPETHOSDVSQVTVWELMTFGAKPDYDGPAREIPDLLEKGERL 947
Db 869 AKTPIKMALESIHFGKYTHOSDVSQVTVWELMTFGAEFYAGLRLAEIPDLLEKGERL 928
Qy 948 POPPICTIDVYIMVKCMIDSECRPRELVSFSSRMARDPQRFVWTONEDLGPASPLD 1007
Db 929 AQOICTIDVYIMVKCMIDENIRPTFKELANETRMARDPPRYLVIKRAS-GFGTP-- 985
Qy 1008 STFYRSLLDDMDGLVDAAEYLVPOQGFPCPDPAAGAGMVHHRHSSSTRSGGDLTL 1067
Db 986 PAAPSPVLTTEL-----QAELEPEL-----DLDL 1011
Qy 1068 GLEPSEE-----EAPRSLAPSEG-----AGSDVFDGLDMG 1099
Db 1012 DLEAEEGLATSLGSLSLPTGTLTRPRGSQSLSPSSGYPMNQSSLGEACLDSAVLGG 1071
Qy 1100 AAKGLQSLPHTDPLQRYSEDPTVPLPSETDGVY---APL-----TC-----SPQP 1143
Db 1072 REQFSRPLSLH-PIPRGR-----PASESEGHVTSGEAELOEKSVVCRSRSRSPRP 1123
Qy 1144 E-----YVNPQDVRPQPSPREGP-----LPAARPAGATLERAKTLSP-GKNGV 1187
Db 1124 RGDAXHYSQRHSLTPTVPLPSPGLEEDNGYVMPDTHLRGASSREGTLLSVGLSSVL 1193
Qy 1188 -----KDVFAFGAVENPEYLTPOGGNAPOPHP 1216
Db 1184 GTEEEDE-----BEYEYMNRRKRGSP-PRPP 1209
```

RESULT 11

EGFR_DROME

ID EGFR_DROME

AC P0412; 061601; Q9W2G0; P81868;

DT 13-AUG-1998 (Rel. 05, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Egfr)

DE (Gurken receptor) (Torpedo protein) (Drosophila relative of ERBB).

```
GN EGFR OR TOP OR C-ERBB OR DER OR CG10079.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS TYPES I AND II).
RX MEDLINE=94350209; PubMed=8070664;
RA Clifford R., Schupbach T.;
RT "Molecular analysis of the Drosophila EGF receptor homolog reveals
RT that several genetically defined classes of alleles cluster in
RL subdomains of the receptor protein.";
RL Genetics 137:531-550(1994).
RN [2]
RP REVISIONS.
RA Clifford R., Schupbach T.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=85124611; PubMed=2982499;
RA Livneh E., Glazer L., Segal D., Schlessinger J., Shilo B.-Z.;
RT "The Drosophila EGF receptor gene homolog: conservation of both
RT hormone binding and kinase domains.";
RL Cell 40:599-607(1985).
RN [4]
RP SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RX STRAIN=Oregon-R; TISSUE=Embryo;
RL MEDLINE=8702474; PubMed=3093080;
RA Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;
RT "Alternative 5' exons and tissue-specific expression of the
RT Drosophila EGF receptor homolog transcripts.";
RL Cell 46:1091-1101(1986).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION ANALYSIS.
RX MEDLINE=99102120; PubMed=9882502;
RA Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;
RT "Several levels of EGF receptor signaling during photoreceptor
RT specification in wild-type, Ellipse, and null mutant Drosophila.";
RL Dev. Biol. 205:129-144(1999).
RN [6]
RP SEQUENCE FROM N.A. (ISOFORM TYPE I).
RX STRAIN=Berkley;
RL MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Anantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclib J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
```

Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
Svirekas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
Wang Z.-Y., Wassaman D.A., Weinstein G.M., Weissenbach J.,
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
"The genome sequence of *Drosophila melanogaster*";
Science 287:2185-2195(2000).
[7]
SEQUENCE OF 959-1078 FROM N.A.
RT STRAIN=Daekwanryeong;
RC MEDLINE=85137938; PubMed=2983232;
RA Wadsworth S.C., Vincent W.S. III, Bilodeau-Wentworth D.;
RX "A *Drosophila* genomic sequence with homology to human epidermal
RT growth factor receptor.";
RL Nature 314:178-180(1995).
RN [8]
SEQUENCE OF 1133-1137; 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION
RP ANALYSIS.
RX MEDLINE=92038942; PubMed=1936959;
RA Raz E., Schejter E.D., Shilo B.Z.;
RX "Intralentic complementation among DER/fib alleles: implications for
RT the mechanism of signal transduction by receptor-tyrosine kinases.";
RL Genetics 129:191-201(1991).
RN [9]
REVIEW.
RX MEDLINE=97248481; PubMed=9084709;
RA Ferrimon N., Perkins L.A.;
RX "There must be 50 ways to rule the signal: the case of the *Drosophila*
RT EGF receptor.";
RL Cell 89:13-16(1997).
RN [-]
FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, CURKEN, VEIN AND ARGOS,
CC WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAP-
CC MAPK PATHWAY. INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS.
CC CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE
CC DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL
CC POLARITIES OF THE OOCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE
CC ESTABLISHMENT OF VENTRAL CELL PATTERNS, MAINTENANCE OF AMNIOGENESIS
CC AND VENTRAL NEUROECTODERMAL CELLS, GERM BAND RETRACTION, CELL FATE
CC SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF
CC CUTICLE.
CC -! CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -! SUBCELLULAR LOCATION: ISOFORM TYPE I AND TYPE II ARE TYPE I
CC MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE
CC PROTEIN.
CC -! ALTERNATIVE PRODUCTS: 3 ISOFORMS; TYPE I (SHOWN HERE), TYPE II AND
CC TYPE III; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -! TISSUE SPECIFICITY: UBICUOUSLY EXPRESSED IN EMBRYOS. IN LARVAE,
CC UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANLAGEN OF
CC TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST
CC EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW, LEVELS REMAIN HIGH
CC IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN
CC POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX
CC AND THORACIC AND ABDOMINAL GANGLIA.
CC -! SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC ENBL; AF052754; AAC08536.1; -
DR ENBL; AF052753; AAC08536.1; JOINED.
DR ENBL; AF052754; AAC08535.1; -
DR ENBL; AF052752; AAC08535.1; JOINED.
DR ENBL; K03054; AAA51462.1; -
DR ENBL; K03417; AAA51460.1; -
DR


```

Db 59 NGSKTPSTAAGVGGCLLVVGLGIGLYLRR-HIVKRTLRLLQERELVEPLTPSGE 117
Qy 712 MPNOAQMRLKETELRKVKVLSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANK 771
Db 118 APNOAHLRLKETEFKVKVLSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANK 177
Qy 772 EILDEAYMAGVGSPPVSRLLGICLTSTVOLVTLMPYGLLDHVRNCRGLSGODLNNW 831
Db 178 EILDEAYMASVDNPHVCRLLGICLTSTVOLVTLMPYGLLDHVRNCRGLSGODLNNW 237
Qy 832 CMQIAGKMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVP 891
Db 238 CVQIAGKMNYLEERLVRDLAARNVLKTPQHVKITDFGLAKLLGADKEYHAEGGKVP 297
Qy 892 IKWMALESILRRRPTHQSDVMSYGVTVWELMTFGKPYDGPAREIPDLLEKGRRLPQPP 951
Db 298 IKWMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGPAREIPDLLEKGRRLPQPP 357
Qy 952 ICTIDVYMWKCMWIDSECPRELVSERFMRDQRFVVIQ-NEDLGPASPLDSTF 1010
Db 358 ICTIDVYMWKCMWIDSECPRELVSERFMRDQRFVVIQ-NEDLGPASPLDSTF 1010
Qy 1011 YRSILLEDDMDGLVDAEYLVPOQGFPCDPAPGAGGMVHRHRSSTRSGGDLTLGLE 1070
Db 418 YRTLEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
Qy 1071 PSEBEAPRSL-----APSEGAGSDVFDGLGMAKGLQLSLPHDPSLPORYEDPTVP 1125
Db 450 -----SRTPLLSLSATSNNSATNCID-----RNGQGHVREDSFVQRYSSDPTGN 495
Qy 1126 LPSET-DGYVAPLTCSPQEVNQDVRPQPPSPREGPLPAARAGATLERAKTLSPGK 1183
Db 496 FLESDIDGFL-----PAPEYVQ--LMPKKPS-----TAMVQ 526
Qy 1184 NGVVKDVP-----AFGAVENPEYLTPOGGAPOPHPPAPAFSDNLYY 1228
Db 527 NQIYNNISLTAISKLPMDSRYSQNSHSTAVDNPEYL-----NTNQSPKLVTFESSPY 578
Qy 1229 WDQ-----DPPE-----RGAPPTFGTPTAENPEYLGIDVP 1260
Db 579 WIOGSHQINLDNPDYQODFLPNETKPNGLLKPAAENPEYLRVAAP 625

RESULT 13
ID ERBB_AVIER STANDARD; PRT; 604 AA.
AC P00535;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbb (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ES4).
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_Taxid=79685;
RN [1]
RC SEQUENCE FROM N.A.
RX MEDLINE=84026539; PubMed=6313229;
RX Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erbb gene of avian erythroblastosis virus is a member of the src
RT gene family."
RL Cell 35:71-78 (1983).
RN [2]
RP SEQUENCE OF 1-152 FROM N.A.
RX MEDLINE=84223957; PubMed=6328658;
RX Debutre B., Henry C., Benaisa M., Biserte G., Claverie J.-M.,
RA Saule S., Martin P., Stehelin D.;
RT "Sequencing the erba gene of avian erythroblastosis virus reveals a
RT new type of oncogene."
RL Science 224:1456-1459 (1984).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein

```

```

CC tyrosine phosphate.
CC -!- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC IN CHICKENS.
CC -!- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; K02006; AAA42394.1; ALT INIT.
CC EMBL; K01216; AAA42400.1; -.
CC
CC DR PIR; A0644; TVYUH.
CC HSP; P11362; IFGK.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR001245; Tyr_pkinase.
CC Pfam; PF00069; pkinase; 1.
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00219; TyrKc; 1.
CC PROSITE; PS00107; PROTEIN KINASE ATP; 1.
CC PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
CC PROSITE; PS00111; PROTEIN KINASE_DOM; 1.
CC Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
CC Glycoprotein; Phosphorylation.
CC
CC DOMAIN 132 399
CC NP_BIND 138 146 ATP (BY SIMILARITY).
CC BINDING 165 165 ATP (BY SIMILARITY).
CC ACT_SITE 257 257 BY SIMILARITY.
CC CONFLICT 29 29 R -> W (IN REF. 2).
CC CONFLICT 140 140 S -> F (IN REF. 2).
CC CONFLICT 146 146 I -> V (IN REF. 2).
CC
CC SQ SEQUENCE 604 AA; 67633 MW; 76EBCDD06745D609 CRC64;
CC
CC Query Match 24.9%; Score 1703; DB 1; Length 604;
CC Best Local Similarity 52.2%; Pred. No. 8.2e-84;
CC Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 16;
CC
Qy 593 CAHYKDPFPCVACPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPEAQ 652
Db 3 CAHPDGFHCVKACPAVGLGENDTL-VRKYADANAVCQLCHPNCRTCKGPGLEGCP--- 58
Qy 653 RASPLTSIVSAVV-GILLVVVLGVVFGILIKRRQKIRKYMRLLOETELVEPLTPSGA 711
Db 59 NGSKTPSTAAGVGGCLLVVGLGIGLYLRR-HIVKRTLRLLQERELVEPLTPSGE 117
Qy 712 MPNOAQMRLKETELRKVKVLSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANK 771
Db 118 APNOAHLRLKETEFKVKVLSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANK 177
Qy 772 EILDEAYMAGVGSPPVSRLLGICLTSTVOLVTLMPYGLLDHVRNCRGLSGODLNNW 831
Db 178 EILDEAYMASVDNPHVCRLLGICLTSTVOLVTLMPYGLLDHVRNCRGLSGODLNNW 237
Qy 832 CMQIAGKMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVP 891
Db 238 CVQIAGKMNYLEERLVRDLAARNVLKTPQHVKITDFGLAKLLGADKEYHAEGGKVP 297
Qy 892 IKWMALESILRRRPTHQSDVMSYGVTVWELMTFGKPYDGPAREIPDLLEKGRRLPQPP 951
Db 298 IKWMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGPAREIPDLLEKGRRLPQPP 357
Qy 952 ICTIDVYMWKCMWIDSECPRELVSERFMRDQRFVVIQ-NEDLGPASPLDSTF 1010
Db 358 ICTIDVYMWKCMWIDSECPRELVSERFMRDQRFVVIQ-NEDLGPASPLDSTF 1010
Qy 1011 YRSILLEDDMDGLVDAEYLVPOQGFPCDPAPGAGGMVHRHRSSTRSGGDLTLGLE 1070
Db 418 YRTLEEDMEDIVDAEYLVPHQGF-----NSPST----- 449

```



```

Db 418 YRTLMEEEDMEDIIVDAEYLVPHQGF-----NSPST----- 449
Qy 1071 PSEEAAPRSL-----APSEGAGSDVFDGLGMAAKGLQLPHTDPSPLQRYSEDPTVP 1125
Db 450 -----SRTPLLSLSLATSNNATNCID-----RNGQGHVREDSEFVQRYSSDPTGN 495
Qy 1126 LPSET--DCYVAPLTCSPQPEYVQDPVROPSPREGPLPAARPAGATLERAKTLPSPK 1183
Db 496 FLEESIDGFL-----PAPEYVQ--LMPKPKSTAM----- 524
Qy 1184 NGVVKDVFAP-----GGAVENPEYLTPOGGAAPQHPHPPAFSPAFD 1224
Db 525 --VQNOIYNFISLTAISKLPMSRYQNSHSTAVDNPEYL-----NTNOSPLAKTVFE 574
Qy 1225 NLYYWDQPPRGAPPSTFKGTPTAENPEY 1254
Db 575 SSPYWIQSNHQ-----INLDNPDY 594

RESULT 14
ID ERBB AVIEU STANDARD; PRT; 540 AA.
AC P11273;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ts167).
OC Viruses; Retroviral viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=103898;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87064458; PubMed=2878364;
RA Choi O.R., Trainor C., Graf T., Beug H., Engel J.D.;
RT "A single amino acid substitution in v-erbB confers a thermolabile
RT phenotype to ts167 avian erythroblastosis virus-transformed erythroid
RT cells.";
RL Mol. Cell. Biol. 6:1751-1759(1986).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M13179; AAA42401.1; -.
CC PIR; A25231; TVFVEB.
CC HSP; P11362; IFGK.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR001245; Tyr_pkinase.
CC Pfam; PF00069; pkinase; 1.
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00219; TyrKc_1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
CC Transferrase; Tyrosine-protein kinase; ATP-binding; Oncogene;
CC Glycoprotein; Phosphorylation.
CC DOMAIN 132 399. PROTEIN KINASE.
CC FT NP_BIND 138 146. ATP (BY SIMILARITY).
CC FT BINDING 165 165. ATP (BY SIMILARITY).
CC FT ACT_SITE 257 257. BY SIMILARITY.
CC FT VARIANT 270 270. H -> D (IN THERMOLABILE V-ERBB).
CC SQ SEQUENCE 540 AA; 60412 MW; 5B53297AA068B65D CRC64;

```

Query Match 23.8%; Score 1630; DB 1; Length 540;
 Best Local Similarity 55.1%; Pred. No. 5.6e-80;

```

Matches 341; Conservative 69; Mismatches 119; Indels 90; Gaps 14;
Qy 593 CAHYKDPFPCVACPSGVKFDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCRAEQ 652
Db 3 CAHFDGPHCVKACPAAGVLGENDTL-VWKYADANAVCOLCHPNCNTRCGCKGPGLEGCP--- 58
Qy 653 RASPLTSIVANV--GILLVVLGVVFGVGLIKRROOKIRKYTMRLLOETELVEPLTPSGA 711
Db 59 NGKSTPSIAAGVVGGLLCLVVGGLGIGLYLRR--HIVKRTLRLLQERELVEPLTPSGE 117
Qy 712 MPNOAQRILKETELRKVKVLGSGAGFTVYKGIWIPGENVKIPIVAIKVIRENTSPKANK 771
Db 118 APNAHURILKETEFKVKVLGFGAGFTVYKGLWIPGEKVTIPIVAIKELREATSPKANK 177
Qy 772 EILDEAYVMAGVGSPPYVSRLLIGICLTSTVOLVLTOLMPYGCLLDHRVNRGRGLSQDLNW 831
Db 178 EILDEAYVMASVDNPHVCRLLGICLTSTVOLITOLMPYGCLLDIYREHKONIGSQYLLNW 237
Qy 832 CMQIAKMSYLEVDVRLVHRDLAARNVLKSPNVHKITDFGLARLLDIDETVHADGGKVP 891
Db 238 CVQIAKGMNYLEERHVMVHRDLAARNVLVKTPOHVKITDFGLAKQLGADEKEYHAEGGKVP 297
Qy 892 IKMALESILFERETHOSDVMSYGVTVWELMTFGAKPYDGIPIABEIPDLLEKGERLPPOP 951
Db 298 IKMALESILERIYTHOSDVMSYGVTVWELMTFGSKPYDGIPIABEISVLEKGERLPPOP 357
Qy 952 ICTIDVYMIWVKWMDSECRPRFRELVSFNSMRDPQRFVVIQ--NEDLGPASPLDSTF 1010
Db 358 ICTIDVYMIWVKWMSDADSRPKRELIAEFKWARDPPRYLVLIQGERMHLPSFTDSKF 417
Qy 1011 YRSILLEDDMDLVDAEEYLVPOGFPFCDPAPAGAGMWHHRSSSTRSGGDLTLGLE 1070
Db 418 YRTLMEEDMEDIIVDAEYLVPHQGF-----NSPST----- 449
Qy 1071 PSEEAAPRSL-----APSEGAGSDVFDGLGMAAKGLQLPHTDPSPLQRYSEDPTVP 1125
Db 450 -----SRTPLLSLSLATSNNATNCIDRNG-----H----- 476
Qy 1126 LPSETDCYVAPLTCSPQPEYVQDPVROPSPREGPLPAARPAGAT--LERAKTLPSPKN 1184
Db 477 -PYREDGFL-----PAPEYVQ--LMPKPKSTAMVQNIYNSLTAISKLPIDSRYQN 527
Qy 1185 GVVKDVFAGGAVENPEYL 1203
Db 528 -----SHSTAVDNPEYL 539

RESULT 15
ID EGFR CHICK STANDARD; PRT; 703 AA.
AC P13387;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER
DE (Fragment)).
DE EGFR.
GN Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88261272; PubMed=3260329;
RA Lax I., Johnson A., Howk R., Sap J., Bellot F., Winkler M.,
RA Ullrich A., Vennstrom B., Schlessinger J., Givol D.;
RT "Chicken epidermal growth factor (EGF) receptor: cDNA cloning,
RT expression in mouse cells, and differential binding of EGF and
RT transforming growth factor alpha.";
RL Mol. Cell. Biol. 8:1970-1978(1988).
CC -!- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:22:34 ; Search time 48.5887 Seconds
(without alignments)
5347.444 Million cell updates/sec

Title: SEQ4-59-73-14
Perfect score: 6847
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGGLDPV 1261

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL 21:.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6202	90.6	1259	6	O18735
2	3101	45.3	1209	11	Q9QX70
3	3072	44.9	1210	11	Q9EP98
4	2729.5	39.9	1137	13	Q9W6F6
5	2683	39.2	1165	13	Q9YH40
6	2263	33.1	1328	13	P79754
7	1959.5	29.2	1433	5	Q9BIH9
8	1775.5	25.9	419	4	Q9UK79
9	1739	25.4	367	11	Q8RX1
10	1720	25.1	729	15	Q86712
11	1718	25.1	567	15	Q86714
12	1697.5	24.8	412	4	Q8WYV0
13	1653.5	24.1	962	15	Q64895
14	1645	24.0	545	15	Q85468
15	1463.5	21.4	655	11	Q9WVF5
16	1447.5	21.1	643	11	Q9ERV6

ALIGNMENTS

RESULT 1

O18735 ID O18735 PRELIMINARY; PRT; 1259 AA.
AC O18735;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Erbb-2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Yokota H.;
RT "CDNA cloning of erbb-2 from canine mammary gland.";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB008451; BAA23127.1;
DR HSSP; P11362; IFGK.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Pfam; PF02757; YLP; 3.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1259 AA; 137989 MW; E37364D49C4ACD46 CRC64;

Q9YLX8 ephydatia f
Q23821 caenorhabdi
Q26566 schistosoma
Q90836 gallus gall
Q9880 rattus norv
Q14256 homo sapien
Q9PSH2 gallus gall
Q923V5 rattus norv
P11776 xiphophorus
Q8SZW1 drosophila
Q9PVZ4 xenopus lae
Q99162 xiphophorus
Q9NJV5 bromphalari
Q9BG66 oryctolagus
Q93457 scophthalmu
Q73798 xenopus lae
Q8UW85 paralichthy
Q8UW86 paralichthy
Q9BUD7 homo sapien
Q8UW84 paralichthy
Q8UW83 paralichthy
Q9QVW4 rattus sp.
Q9YGH8 scophthalmu
Q9U5A8 bombyx mori
Q9VD94 drosophila
Q91YM0 mus musculu
Q96L35 homo sapien
Q99MR2 mus musculu
Q07912 homo sapien

Query Match		90.6%; Score 6202; DB 6; Length 1259;
Best Local Similarity		90.7%; Pred. No. 0;
Matches 1148; Conservative 42; Mismatches 64; Indels 12; Gaps 4;		
Qy	1	MELAAALRWGLLLALLPPGAASVCTCTDKMLRLPASPETHDMLRLHYGCGVQGNL 60
Db	1	MELAAALRWGLLLALLPPGAASVCTCTDKMLRLPASPETHDMLRLHYGCGVQGNL 60
Qy	61	ELTYLPTNASLFLQDIOEVEFNFTSFVLRVKSASHLEQRILVRGTLQDFEDNAL 120
Db	61	ELTYLPANASLFLQDIOEVEFNFTSFVLRVKSASHLEQRILVRGTLQDFEDNAL 114
Qy	121	AVLNDGDDPLNNTPTVTGASPGELRELQRLSITEILKGGVLIQRPOLCYQDTILWKDIFH 180
Db	115	AVLNDGDDPLNNTPTVTGASPGELRELQRLSITEILKGGVLIQRPOLCHQDTIILWKDVFH 174
Qy	181	KNNQALALTLIDNRSRACHPCSPMCKGSRCKGSESSDCQSLTRTVACGGCARCKGPLPTD 240
Db	175	KNNQALALTLIDNRSRACHPCSPMCKGSRCKGSESSDCQSLTRTVACGGCARCKGPOPTD 234
Qy	241	CCHQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESHPNPEGRYTFGAS 300
Db	235	CCHQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESHPNPEGRYTFGAS 294
Qy	301	CVTACPNYLTSDVGSCTLVCPHNOEYVTAEDGTQRCCKSPCARVCYGLGMEHLREVR 360
Db	295	CVTSCPNYLTSDVGSCTLVCPHNOEYVTAEDGTQRCCKSPCARVCYGLGMEHLREVR 354
Qy	361	AVTSANIOEAGCKKIFGSLAFLESFDGDPASNTAPLQPEQLQVFLEBITGYLYISA 420
Db	355	AVTSANIOEAGCKKIFGSLAFLESFDGDPASNTAPLQPEQLRVFEALEBITGYLYISA 414
Qy	421	WPDLSPLDSVFQNLQVIRGRILHNGAYSITLQGLISWLGRLSRELGSGLALIHNNTHL 480
Db	415	WPDLSPLDSVFQNLQVIRGRILHNGAYSITLQGLISWLGRLSRELGSGLALIHNNRL 474
Qy	481	CFVHTVPMDQLFRNPHQALLHTANRPEDECVEGEGGLACHOLCARGHCWGPGTQCVCNQSF 540
Db	475	CFVHTVPMDQLFRNPHQALLHTANRPEDECVEGEGGLACHOLCARGHCWGPGTQCVCNQSF 533
Qy	541	LRGQECVEECRVQLGPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDDPP 600
Db	534	LRGQECVEECRVQLGPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDDPP 593
Qy	601	FCVARCPGVKPDLSYMPDIWFPDEBEGACQCPINCTHSCVDLDKGCAPQORASPLTSI 660
Db	594	FCVARCPGVKPDLSYMPDIWFPDEBEGACQCPINCTHSCVDLDKGCAPQORASPLTSI 653
Qy	661	VSADVGLLVVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAQMRI 720
Db	654	IAAVVGILLAVVGLVGLILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAQMRI 713
Qy	721	LKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVIRENTSPRANKEILLDEAYVM 780
Db	714	LKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVIRENTSPRANKEILLDEAYVM 773
Qy	781	AGVGSPPYVRLIGICLTSTVQLMPYGLLDHVRNRRGLSGQDLLNCKMCIAGMS 840
Db	774	AGVGSPPYVRLIGICLTSTVQLMPYGLLDHVRNRRGLSGQDLLNCKMCIAGMS 833
Qy	841	YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKKNWALESI 900
Db	834	YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKKNWALESI 893
Qy	901	LRRRFTHQSDVMSYGVTVWELMTGAKPYDGIPIAREIPDLLEKGERLPQPPICTIDVYMI 960
Db	894	PPRRFTHQSDVMSYGVTVWELMTGAKPYDGIPIAREIPDLLEKGERLPQPPICTIDVYMI 953
Qy	961	MVKCMWIDSECRPRELVSFSRWARDPQRFVJQNEDLGPASPLDSTFYRSLLEDDDM 1020
Db	954	MVKCMWIDSECRPRELVSFSRWARDPQRFVJQNEDLGPASPLDSTFYRSLLEDDDM 1013
Qy	1021	GDVDAEYLVFQOQFFCPEPTAGAGGMVHRRSSSTRSGGDLITLGLPSEBEPKSP 1080
Db	1014	GDVDAEYLVFQOQFFCPEPTAGAGGMVHRRSSSTRSGGDLITLGLPSEBEPKSP 1073
Qy	1081	LAPSEGAGSVDFGDLGMAAGKGLQSLPHTDPPQLQRYSEDPTVPLPSETDGYVAPLTCS 1140
Db	1074	LAPSEGAGSVDFGDLGMAAGKGLQSLPHTDPPQLQRYSEDPTVPLPSETDGYVAPLTCS 1133
Qy	1141	POEYVNOQDVROPPSPREGPLPAARPAGATLER-----AKTILSPCKNGVVKDVFAGG 1195
Db	1134	POEYVNOQDVROPPSPREGPLPAARPAGATLERPKTILSPCKNGVVKDVFAGG 1193
Qy	1196	AVENPEYLTPOGGAAPQHPHPPAFSPAFNDLYWDQDPPRGAPPSTFKGTPTAENPEYL 1255
Db	1194	AVENPEYLAIRGRAAPQHPHPPAFSPAFNDLYWDQDPPRGAPPSTFKGTPTAENPEYL 1253
Qy	1256	GLDVPV 1261
Db	1254	GLDVPV 1259
RESULT 2		
ID	Q9QX70	PRELIMINARY; PRT; 1209 AA.
AC	Q9QX70;	
DT	01-MAY-2000 (TrEMBLrel. 13, Created)	
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)	
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)	
DE	Epidermal growth factor receptor.	
GN	EGFR.	
OS	Rattus norvegicus (Rat).	
OC	Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
OX	NCBI_TaxID=10116;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=FISHER; TISSUE=LIVER;	
RC	MEDLINE=90258888; PubMed=2342466;	
RA	Petch L.A., Harris J., Raymond V.W., Blasband A.J., Lee D.C.,	
RA	Earp H.S.;	
RT	"A truncated, secreted form of the epidermal growth factor receptor is	
RT	encoded by an alternatively spliced transcript in normal rat tissue."	
RL	Mol. Cell. Biol. 10:2973-2982(1990).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=FISHER; TISSUE=LIVER;	
RA	Petch L.A.;	
RL	Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.	
RN	[3]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=FISHER; TISSUE=LIVER;	
RA	Guttridge K., Dawson T.L., Earp H.S.;	
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; M37394; AAF14008.1; -.	
DR	HSP; P11362; IFGK.	
DR	InterPro; IPR000494; EGFR_L domain.	
DR	InterPro; IPR000719; Euk_pkinase.	
DR	InterPro; IPR002174; Furin-like.	
DR	InterPro; IPR001245; Tyr_pkinase.	
DR	Pfam; PF00757; Furin-like; 1.	
DR	Pfam; PF00069; pkinase; 1.	
DR	Pfam; PF01030; Recep_L domain; 2.	
DR	PRINTS; PR00109; TYRKINASE.	
DR	ProDom; PD000001; Euk_pkinase; 1.	
DR	SMART; SM00261; FU; 3.	
DR	SMART; SM00219; Tyrc; 1.	
DR	PROSITE; PS00107; PROTEIN KINASE ATP; 1.	
DR	PROSITE; PS00011; PROTEIN KINASE DOM; 1.	
DR	PROSITE; PS00109; PROTEIN KINASE_TYR; 1.	
KW	ATP-binding; Receptor; Transferase; Tyrosine-protein kinase.	
SQ	SEQUENCE 1209 AA; 134891 MW; 96FEF7F6CC1B7773 CRC64;	
Query Match		45.3%; Score 3101; DB 11; Length 1209;
Best Local Similarity		49.4%; Pred. No. 5.9e-226;

DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00011; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
KW ATP-binding; Receptor; Transferase.
SQ SEQUENCE 1210 AA; 134840 MW; 62CD021C9D3E2E18 CRC64;

Query Match 44.9%; Score 3072; DB 11; Length 1210;
Best Local Similarity 48.9%; Pred. No. 9.3e-224; Indels 118; Gaps 25;
Matches 625; Conservative 171; Mismatches 365;

Qy 11 LLLALLPGAA--STQVCTGTDMLRLPASPETHLDMLRLHYLQCGQVQVQGNLELYLPTN 68
Db 14 LLLTALCAAGGALBEKKVCOGTSNLTQLGTEDHFLSLQRYNNCEVVLGNLEITYVQRN 73

Qy 69 ASLSFLQDIQVQ--FNNFTVSFWLRVPKVSASHLEQRIRIVRGQLPFDNEDYVALAVLNGD 127
Db 74 YDLSFLKTIQEVAGYVLIALTVERIPL-----ENLQIRGNALYENTYVALAILSN-- 124

Qy 128 PLNNTTPVTGASPGGLRELQLSLEILKGVLIQRPOLCYQDTILWKDI---FHKN 183
Db 125 -----YGTNRGLRELPMRNQLEILIGAVRFSNNPILCNMDTIQWRDIVQVFNW 176

Qy 184 QLALTLIDNRSRACHPCSPMKGSRGWESSEDCQSLTRTVACAGCA--RCKGPLPTDCC 242
Db 177 SMDL-----QSHPPSCPCKDPCSPNGSCWGGGEENCQKLTIIICAQCSHRCGRSPSDCC 232

Qy 243 HEQCAAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTPFSMNPENEGRYTFGASCV 302
Db 233 HNQCAAGCTGPRESDCLVQKQFQDEATCKTCTPPLMLYNPTTYQMDVNPGEKSYFGATCV 292

Qy 303 TACPNYLSLDVSGCTLVPLHNOEVAEDTCORCEKSKPCARVCYGLGMEHLREVRVAV 362
Db 293 KKCPRNVYVTDHSGVCRACGPDYEV--BEDGIRCKKCDGCRKVCNGIGIGEFPKTL 351

Qy 363 TSANIQFAGCKTIFGSLAFLEPSFDGDPASNTAPLQPEQLQVPLEETITGYLYISAMP 422
Db 352 NATNKHFKYKTAISGDLHILPVAFKGSFRTPLDPLPRELEILKTVKEITGFLLIQAMP 411

Qy 423 DSLPDLVFQNLQVIRGRILHNGAYSLTQGLGTSWLSRLSRLRELGLALIHNTHLCP 482
Db 412 DNWTDLHAFENLEIRGTQKQGFSLAVVGLNITSLGRLSLEISDGDVLIISGNRLCY 471

Qy 483 VHTVPWOLFERNHQALLHTANREDEECVSGGLACHOLCARGHCGPCPTOCVNCQFLR 542
Db 472 ANTINWKLFTPNQTKIMNRAEKDCKAVHNCVNLCSSEGCWGBEPDPCVSCQVSR 531

Qy 543 QGECVECRVLQGLPREYVVARHCLPCHPECPQNGSVTCFGEADQCVCAHYKDPFPC 602
Db 532 GRECEKCNILLEGPREFVENSECICQHPCLPQAMNITCTGRGPDNCICQAHYIDGPHC 591

Qy 603 VARCPGVKPLSYMPIWKFPPDEGACQPCPINCTHSCVDLDDKGPQABQASPLTSIVS 662
Db 592 VKTCPAGIMGENTL--VMKYADANNVCHLCHANCTYGCAGPGLQCEVWPSPGKIPSIAT 650

Qy 663 AVGGLLVVGLGVFGI--LIERQOKIRKYTWRLLOETELVEPLTPSGAMPNQAORIL 721
Db 651 GIVGGLFIVV--VALGIGLFWRRRHIVKRTLRLLQERLEVEPLTPSGEAPNOAHLRIL 709

Qy 722 KETELRKVKVLGSAFGVYKGIWIPDGENVKIPVAIKVLENTSPRANKELDEAYMA 781
Db 710 KETEFKKIVLGSAGFVYKGLMPEGEKVIKVAIKELREATSPRANKELDEAYMA 769

Qy 782 GVGSPYVRLGICLTSTVQLVTLQMPYGCILLDHHVRENRLGSDLLNWCWQIAKMSY 841
Db 770 SVDNPHVCRLLGICLTSTVQLITQMPYGCILLDYVREHKDNGISQYLLNWCVQIAKGMN 829

Qy 842 LEDVRLVHRDLAARNLVKSNHVKITDFGLARLLDIDETVHADGGKVPKKNWALSIL 901
Db 830 LEDRRLVHRDLAARNLVKTPHVKITDFGLAKLGAEEKYHAEKGVKPKKNWALSIL 889

Qy 902 RRRRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIM 961
Db 902 RRRRTHQSDVMSYGVTVWELMTFGSKPYDGIIPASDISILEKGERLPQPPICTIDVYIM 949

Qy 962 VKCWMIDSECRPRELVSFEFSRMARDPQRFVVIQ--NEDLGPASPLDSTFYRSLLEDDDM 1020
Db 950 VKCWMIDADSRPKRPRELILFEFSKMAADPQRYLVIOGDERMHLFSPDTSNRYALMDEDM 1009

Qy 1021 GDLVDAEYLVPOQGFPCDPAPFAGAGGVHHRSSSTRSGGDLTLGLEPSEEA PRSP 1080
Db 1010 EDVVDADEVLTPOQGF-----NSPST-----SRT 1035

Qy 1081 LAPSEGAGSVDFDGLGMGAAGKQLSLPHTDPSFLQRYSEDPVPLPSET--DGYVAPLT 1138
Db 1036 LLSLSLATS-----NSTVACINRNGSCRVKEDAFQRYSSDPTGAVTEDNIDDAFL---- 1087

Qy 1139 CSQPEVNVQPDYRPPQPPSPREGPLPAARPAAGATLERAKTLSPGKGVKDVAFAGAVE 1198
Db 1088 --VPVEVYVQ--SVPKRPAGSVQNPVYHNPQLHP-----APGRDLHYQN--PHSNAV 1134

Qy 1199 NPEYL--TPQGAAPQHPPPPAPFDNLYWYDQ-----DP-----PERGAPPS 1241
Db 1135 NPEYLNTAQ-----PTCLSSGFSNPAWIKGSHQMSLDNDPDYQODFPFKETKPNG 1185

Qy 1242 TFKGTPTAENPEYLGIDVP 1260
Db 1186 IFKG-PTAENAEYLRVAPP 1203

RESULT 4
Q9W6F6 PRELIMINARY; PRT; 1137 AA.
AC Q9W6F6;
DT 01-NOV-1999 (T-EMBLrel. 12, Created)
DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)
DT 01-JUN-2000 (T-EMBLrel. 21, Last annotation update)
DE Receptor tyrosine kinase (fragment).
GN ERB4.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=HINDERAIN;
RX MEDLINE=9263203; PubMed=10328884;
FA Dixon M., Lumsden A.;
RT "Distribution of neuregulin-1 (nrg1) and erbB4 transcripts in embryonic chick hindbrain."
RL Mol. Cell. Neurosci. 13:237-258 (1999).
DR EMBL; AF121963; AAD31764.1; -;
DR HSP; P11362; IFGK.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001368; TNFR_c6.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004015; YLP_motif.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L domain; 1.
DR Pfam; PF02757; YLP; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TYRK; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00011; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
DR PROSITE; PS00652; TNFR_NGR_1; UNKNOWN_1.
KW Kinase; Tyrosine-protein kinase.
FT NON TER 1
SEQUENCE 1137 AA; 127927 MW; 4D616436F87DC84F CRC64;

Query Match		39.9%; Score 2729.5; DB 13; Length 1137;
Best Local Similarity		47.3%; Pred. No. 7.9e-198;
Matches		539; Conservative 169; Mismatches 354; Indels 77; Gaps 24;
Qy	167	LCYODTLKWDIEHKNNQALTLTDNRSRACHPCSPWCKSGSRGWGESSEDCQSILTRVC 226
Db	3	LCFADTHIQDIIVRNWASNFITLPTNGSSGGRCHKSCCTG-RCWGPFTENHCQILTKVC 61
Qy	227	AGGC-ARCKGPLPTDCCHQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVYNTYDTF 285
Db	62	AEQCDGRCYGPVSDCHRECAGCGSPKDTDFACMNFDSGACVTCQPTQFVFNPTTF 121
Qy	286	ESMPNPGRYTFGACVTCACPNYLSTDVSGCTLVCPLNQEVTAEDQTQCEKCRPCA 345
Db	122	QLEHNHNAITYGAFCKVCKPHNFV-VDSSSVCACPSKMEV-EENGICKKCKCTDICI 179
Qy	346	RVCYGLGMEHREVRVTSANIOFAGCKIFGSLAFLPESFDGDPASNTAPLOEQOV 405
Db	180	KACDGTGSLVSAQTVDSSNIDFEINCKINGNLIFLVTHGDPYHTIAAINPERKNI 239
Qy	406	FETLEEITGYLYISAWPDSLPLSFQNLQVIRGIRLHNGAYSLTQLGLGSLWGLSLR 465
Db	240	FOTVREITGYLNIQSPENMTDFVSNLVTIGGRALYSGLSLLLKQGGITSLQFQSLK 299
Qy	466	ELGSLALIHNTLHCFVHTVPWOLFRNPHQALLHTANRPEDCEVGEGLACHQLCARGH 525
Db	300	QISAGNIYITNSNLCTYHTWTSLSFSTPSQKTIHRNKAENCTADGMVNCLECSDG 359
Qy	526	CWGPPTQCVNCSOFLRGQECVEECVRLQGLPREVNARHGLPCHPEQCP-QNGSVTCFG 584
Db	360	CWGPDPDCLCKFRIGRTCIESCNLVDGFRFANGSVCEWQDQCEKEDNMNITCYG 419
Qy	585	PEADQCACAHYKDPFFCVCARPCSGVKPDLSPYMIWPKFPDEGACQCPINCTHSCVDLD 644
Db	420	PGPDHCKCFHFKDGPCNCEKCPDGLQANSF-IFKYADEDRCHPCNCTQCGRPA 477
Qy	645	DKGC-----PABQRASPLTSIVSAVV-GILLVVLGVVFGILIKRQOKIRKYT 692
Db	478	SHDCIYPTWTRQSTLPQAR-TPL--IAAGVIGLFIIVIMGLTFAVYVRRKSIK-KKRA 533
Qy	693	MRLLOETLEVEPLTPSGAMPNQOMRILKETELRKVKVLGSGAGTGYKGIWTPDGENV 752
Db	534	LRREL-ETLEVEPLTPSGATPNQALRLKETELRKVKVLGSGAGTGYKGIWPEGETV 592
Qy	753	KIPVAIKVLRNTPSKANKELIDRAYVMAGVSPVSRLLIGLICTSTVOLVTLQMPYGL 812
Db	593	KIPVAIKILNETTGPKANVEFMDALINASMDHPLVLLGVLSPITQLVTLMPHGL 652
Qy	813	LDHVRENRLGSQLLNWCQIAKMSYLEEDVRLVHRDLAARNVVKSPNHVKITDFGL 872
Db	653	LDYVHEKDNITGSQLLNWCQIAKGMVYLBERRLVHRDLAARNVVKSPNHVKITDFGL 712
Qy	873	ARLDIDETEYHAGGKVPKWMALLESILRRFTQSDVWSYGYVTWELMTFGAKPYDGI 932
Db	713	ARLLEGDKEYNAGDGKMPIKWMALLECIFYKFKFTQSDVWSYGYVTWELMTFGKPYDGI 772
Qy	933	PARBIPDLLEKGERLPQPICTIDVYIMVWKCMIDSECRPRFRELVESEFMRMARDPORF 992
Db	773	PTREIPDLLEKGERLPQPICTIDVYVMVWKCMIDADSRRFKFELAAEFSEMRMARDQRY 832
Qy	993	VVIQNE-LGPASPLDSTFYRSLLEDDMDGLVDAEYLVPOQGFCCPDPAAGGMVHH 1051
Db	833	LVIQDDRMKLPSPNDSKFPQNLDEEDLEMDMAEYLVF-QAFNIPPIYTSRTIDS 891
Qy	1052	RHRSSSTRSGGDLTLGLFEPSEERAPS--PLAP-SEGAGSDVDGDLGMAAGLQSLP 1108
Db	892	NRNOFVVRDGGYAAEQGV-PMPYRAGCIIIEAPVQAQATAEIPEDTCCNGLRKQVATL 950
Qy	1109	THDPSPLQRYSEDPTVPLPS-----ETDGYVAPLTCSPQEVNPDVBPQPPSPREG 1161
Db	951	AKEDSSQRYASDVTVPFPERVIRGELDEDEGWTMPMDKPTDILNVEENPFVSRKNG 1010
Qy	1162	PLPAA-RPAGATLERAKTSLSPKNGVKDVF-----AFGGAVENPEYLTPOGGAAPQ 1212
Db	1011	DLQAVDNPYHN-----APNGQPKAEDEYVNEPLYLNTFANTLENAEYL-----K 1055
Qy	1213	PHPPAPSPAFNDLYYWDQPPPERGA--PPSTFKGTPT-----AENPEYL 1255
Db	1056	NNLPEKAKKAFNDPDYWNHSLPFRSTLQHPDYLOEYSTKYFYKQNGRIRPIVAENPEYL 1114
RESULT 5		
Q9YH40		
ID	O9YH40	PRELIMINARY; PRT; 1165 AA.
AC	O9YH40	
DT	01-MAY-1999	(TrEMBLrel. 10, Created)
DT	01-OCT-2000	(TrEMBLrel. 15, Last sequence update)
DT	01-JUN-2002	(TrEMBLrel. 21, Last annotation update)
DE	Receptor tyrosine kinase proto-oncogene.	
GN	XMRK	
OS	Xiphophorus xiphidium.	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;	
OC	Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;	
OC	Cyprinodontiformes; Poeciliidae; Xiphophorus.	
OX	NCBI_TaxID=8086;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=RIO PURIFICATION;	
RC	MEDLINE=98241172; PubMed=982016;	
RA	Dimitrijevic N., Winkler C., Wellbrock C., Gomez A., Duschl J.,	
RA	Altschmid J., Scharl M.	
RT	"Activation of the Xmrk proto-oncogene of Xiphophorus by	
RT	overexpression and mutational alterations."	
RL	Oncogene 16:1681-1690(1998).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=RIO PURIFICATION;	
RC	Scharl M.;	
RL	Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; U53471; AAD10500.2; -.	
DR	HSP; P11362; 1FGK.	
DR	InterPro; IPR000345; CytC heme bind.	
DR	InterPro; IPR000494; EGFR_L domain.	
DR	InterPro; IPR000719; Euk_pkinase.	
DR	InterPro; IPR002174; Furin-like.	
DR	InterPro; IPR001899; Gram_pos_anchor.	
DR	InterPro; IPR001245; Tyr_pkinase.	
DR	Pfam; PF00757; Furin-like; 1.	
DR	Pfam; PF00069; pkinase; 1.	
DR	Pfam; PF01030; Recep_L domain; 2.	
DR	PRINTS; PR00109; TYRKINASE.	
DR	ProDom; PD000001; Euk_pkinase; 1.	
DR	SMART; SM00261; FU; 3.	
DR	SMART; SM00219; TyrKc; 1.	
DR	PROSITE; PS00190; CYTOCHROME C; UNKNOWN 1.	
DR	PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN 1.	
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.	
DR	PROSITE; PS50011; PROTEIN_KINASE_DOM; 2.	
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.	
KW	ATP-binding; Kinase; Transferase; Tyrosine-protein kinase.	
SQ	SEQUENCE 1165 AA; 129614 MW; 7F7E38D8771A74E CRC64;	
Query Match		39.2%; Score 2683; DB 13; Length 1165;
Best Local Similarity		44.9%; Pred. No. 2.7e-194;
Matches		577; Conservative 164; Mismatches 385; Indels 158; Gaps 32;
Qy	1	MELAAALCRWGLLALLPPG-AAST-----QVCTGDMKRLRPASPETHDMLRHLVQCGOV 55
Db	4	LELLEL-----LALLLSIGRCCSTDPDRKVCQGTSNQMTM---LDNHLYLKMKNYSGCNV 56
Qy	56	VQGNLELTYPNTASLFLODIOEV-----QFNFTVSWFLRPVKPSASHLEQLRLIVR 109
Db	57	VLENLEITYTQHNOQLSFLQSIQEVGGYVLIAMNEVST-----IPLVN-----LKLIR 104
Qy	110	GTQLFEDNYALAVLDNGDPLNNTTPTVTGASPGGLRELQLRSLTEILKGGVLIQNPOLCY 169


```
Db 105 GONLYEGNFTLLVMSNYOK-NPSSP--DVYQVGLKQLSLNLTSLSGVKVSHNPLLCN 161
Qy 170 QDTILWKDIFHNQALATLIDTNRSRACHPCSPMCKGSRGWGSESDCQSLTRTVGAGG 229
Db 162 VETINWMDIVDTSNTPTNLIPAFERQOCQKDCPGCVNGSCWAFQPGHCQKFTKLLCAEQ 221
Qy 230 C-ARCKGPLPTDCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFSM 288
Db 222 CNRRCRGPKPIDCCNEHCAGCTGPRATDCLACRDFNDGCTKCTCPPPKIYDIVSHQV 281
Qy 289 PNPEGRYTFGASCVTACPYNLSTDVGSCITLVCPHLNQEVTAEDQTCERCKSPCARVC 348
Db 282 DNPNIKYTFGAACVKECFSNVWTE-GACVRSAGMLEVD-ENGKRSCKPCDGVCPKVC 339
Qy 349 YGLGMEHLREVRANTSNIQFAGCKKIFGSLAFIPESFDGDPASNTAPLOPEQLVPET 408
Db 340 DGIGTIGLSNTIAVNSTNIGSFNSCTKINGDIILNRNSFEGDPHYKIGPMDPEHLNLT 399
Qy 409 LEEITGYLYISAWPDLPLDSVFONQLOVIRGRIHLNGAYS-LTQGLGIGMLGRSLREL 467
Db 400 VKEITGYLYIWMWENMTSLSFVQNLRIIRGRTTFRSGFSFVVQVSHQWGLGSLSKEV 459
Qy 468 GSGIALIHNTLCFVHTVPDQLFRNPQHALLHTANRPEDECYGEGLACHQLCARGHCW 527
Db 460 SAGNVILKNTPLQRYASTINRRUFRSDEQSIYDART-----ENQTCNECSDGCW 512
Qy 528 GPGPTQVCNCSQFLRGQCEVCEBVLQGLPREYVNRHCLCHPECPQNGSVTCFGPEA 587
Db 513 GPGPTMVCSLHVRDGRGCVASCNLLQGEPPREAQVDGRVCQCHOECLVQTDLSLTCYGGP 572
Qy 588 DQCVACAHYKPPFCFVARCPGSKVPDLSPYMPFWKPFDEEGACQPCINCTHSCVDLDDKG 647
Db 573 ANCSKAIHQDQPCIPRCPHMLGDGTL-IWKYADKMGQCPQCHQNTQCGSGPGLSG 631
Qy 648 CPAEQRASPLTSIVSAVYGILLVVLVGVFGLIKRRQOKIRKYMRLLELQETLVEPLT 707
Db 632 CRGD-IVSHSLAVGLVSGLLITVALLIVLLRRRIK-RKRTIRLLQEKELVEPLT 689
Qy 708 PSGAMPNQAQWRIKTELRLKVKVLGSGAFGTVYGIWIPGENVKIIPVAIKVLRNTSP 767
Db 690 PSGAPNQAFLRIKTELRFKDRVLGSGAFGTVYKGLWNPDPENIRIPVAIKVLRNTSP 749
Qy 768 KANKEILDEAYMAGVGSPPYVRLIGLICTSVOLATOLMPYGCILDHVRNRLGSGD 827
Db 750 KQNGEVLDEAYMASVDHPHVCRLGICLTSAVQLVTQMPYGCILDHVRQERICQW 809
Qy 828 LLNMCQIAKMSYLEVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADG 887
Db 810 LLNMCQIAKGMNYLEERHLVHRDLAARNVLKSPNHVKITDFGLSKLLTADEKEYQAHG 869
Qy 888 GKVPKIMWALESLRRRTTHOSDQVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGBRL 947
Db 870 GKVPKIMWALESLIOWTTHOSDQVWSYGVTVWELMTFGSKPYDGIIPAKIEASVLNGERL 929
Qy 948 POPPCTTIDVMYIMVKWIDSECPREFELVSEFSRMARDPQRFVVIQONEDLGPASPLD 1007
Db 930 POPPCTTIEVMYILKCMWIDPSRPRELVEFSQARDPSRYLVLTQG--NLPSPSD 986
Qy 1008 STEFVRLSLDEDDMDGLVDAEYLVPOQGFPCPDAPGAGGMVHRHRSSTRSGGDLTL 1067
Db 987 RRLFSRLSSDD--DVVDADVLL-----RYKRIN-RQGS----- 1018
Qy 1068 GLEPSEEARPSPLAPSGAGSDVPDGLGMAKGLQSLPTHDPSPLOQRYSEDPV-PL 1126
Db 1019 -----EPCIPPNH-----FVRENSIALRYSIDPTQNAL 1047
Qy 1127 PSETDGYVAPLTCSPQPVYVNPQVVRPQ-----PSPRE-----GLP-AARPAGATL 1173
Db 1048 EKDLGDH-----EYVNPQGETSSRLSDIYNPNYEDTDCGWGVSUSSQAEATNF 1097
Qy 1174 ERAKTLSPGKNGVKNVDVFAFGAVENPEYLTPOGGAAPQPPPPAFSPAFNLYYWDQD 1233
```

```
Db 1098 SRPEYLTNTQNSL---PLVSSGSMDDPDY---QAG-----YQAAF-----L 1132
Qy 1234 PERGAPPSTFKGTPTAENPEVIGL 1257
Db 1133 POTGALTGNGMFUPAAENLEYLGL 1156

RESULT 6
P79754
ID P79754 PRELIMINARY; PRT; 1328 AA.
AC P79754;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE ErBB3.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Teleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Takifugu.
OX NCBI_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99177347; PubMed=10077531;
RA Gellner K., Brenner S.;
RT "Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu
rubripes.";
RL Genome Res. 9:251-258(1999).
DR EMBL; AF056116; AAC34391.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyK; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00011; PROTEIN KINASE_DOM; 1.
DR ATP-binding; Transferase.
SQ SEQUENCE 1328 AA; 148613 MW; A333039258B647E9 CRC64;

Query Match 33.1%; Score 2263; DB 13; Length 1328;
Best Local Similarity 39.9%; Pred. No. 2.3e-162;
Matches 517; Conservative 157; Mismatches 418; Indels 204; Gaps 34;

Qy 9 WGLLLALLPP--GAASTQ---VCTGTDMLRLPASPETHLMLRLHYQGCVVQGNLEI 62
Db 4 WRLTLMCVASRLRAASQTQEAFCPTQNGLSSTGSGENQVNLNKDKYKGCETIMGNLEI 63
Qy 63 TYLPTNASLFLQDIOEQVFNNTVSEFWLRVPKVSASHLEQ-----RLRIVRGTLQFEDNY 118
Db 64 TQIESNWFSLTKTIREV-----TGVVL---IAMNHQEIPLGLRVIRGNSLYERRF 113
Qy 119 ALAVLDNGDPLNNTPTVYGASPGGLRELQLRLSLTEILKGGVLIORNPOLCYQDTILWKDI 178
Db 114 ALSVFLN-----YPKDG--PSGLNQLGLMLNLTLEILDGGVQIINNKYLRYPVWYWRDI 164
Qy 179 FHKNQLALTLIDTNRSRACHPCSPMCKGSRGWGSESDCQSLTRTVGAGG-ARCKGPL 237
Db 165 I-RNNDAPIEIQFNGERGVCVH---KSC-GNYCWPFGDKQQQLTKTVCAPOCNDRCFGTS 219
Qy 238 PTDCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFSMNPPEGRYTF 297
Db 220 PRDCCHECAAGCKGPLDTCFACRLFNDGACVPQCPQTLIYNKQTFQMETNPAKYOY 279
Qy 298 GASCVTAAPNYLSTDVGSCITLVCPHLNQEVTAEDGTQ-CEKSKPCARVCYGLGMEHL 356
Db 280 GSIQVSCQPTHFV-VDGSSCVSVCPDPKMEV--ERGSQROCELCSGLCPKVCETGAE-- 334
```



```
Qy 489 DQFRNPHQALLHTANRPEDECVEGLACHQLCARGHGWGPPTQCVNCSQFLRQECVE 548
Db 444 SEIKKSDHEVMVQKRNATECHEGMECSQCSKAGCWGKPGQCLECKNVKYGKCLD 503
Qy 549 ECRVLQGLPRY-VNARHCLPCHPECPQNGSVTCFGEADQCACAHYKDPFPCVACP 607
Db 504 SKC---SLPRLYSVDSKTCGCHQCKD-----FCYGPNEIDCGSCMNVKDGRCVACP 555
Qy 608 -----SGVKPDLVYMPIKFPD-- 624
Db 556 TTKHAMNGTCINCHKTCVCGCRPRDTIAPDGCISCDKAIIGSDAKIERCLMKDSCPDGY 615
Qy 625 -----EEG----- 627
Db 616 YSDVLOEGLKQLSGKAVCRKCHPRCKKCTGYGFHEQFQCECTGYKKGCEDECPQD 675
Qy 628 -----ACQPCINCT-----HSCVDL-----DD-----KGCBAEQ--- 652
Db 676 FYANETRILCLPCHQECRGCHGLGDHHECRNLKLFEGDPYDNNATTFTCVSNCPASHPYK 735
Qy 653 -----RASPLTSIVSAVVGILLVVVLGVFGI---LIKRQOK 687
Db 736 RFPQAGKIGYCSADSNQSGRLIEPQTVKIVMGVSNALILLCVFGIAVLFPSRHKN 795
Qy 688 IRKYMTRRLLOETELVLEPLTPSGAMPNOAQWRIKTELKRVKVLGSGAFVYKGIWIP 747
Db 796 KDAVMTMALAGCEDSEPLRSNVGNLTKRIIEAIEIRGGVGLMGAFGRVFGVWMP 855
Qy 748 DGENVKIPVAIKVIRENTSPKANKEILDEAYMAGVSGPYVSRLLGICLTSTVOLVQLM 807
Db 856 EGESVKIPVAIKVLMESGSSSEKFELEAYIMASVEHPNLLKLLAVCMTSQMMILITQLM 915
Qy 808 PYGCLLDHVRNCRGLCSODLNNCMOIAKMSVLEOVLRVHRLDAAARNVLKSPNHVKI 867
Db 916 PLGCLLDYVRNKKDKIKSALNNWSTQIAREMYLEERRLVHRLDAAARNVLQVTPSCVKI 975
Qy 868 TDFGLARLLDDETEYIADGGKVPKIMWALSESILRRFTTHOSVMSYGVTVWELMTFGAK 927
Db 976 TVFGLAKLLDFSDYRAAGGKMPKMLALECIRHRVFTSKSDVWAFGIITWELLTYGAR 1035
Qy 928 PYDGIAREIPDLLEKGERLPPOPICTIDVIMVWKMDISECRPRPRELVSEFSMAR 987
Db 1036 PYENVPKDVPELIEIGHKLPQPDICSLDVYCIILSCWLDADARPTFKLAETFAEKAR 1095
Qy 988 DPQRFVVIQNEIDLGPASPLDSTFYRSILLEDDMDGLV----- 1024
Db 1096 DPGRYLMI-----PGDKFMLPLPSYTNODEKDLIRTLAPVAMAAAAAAGASNV 1146
Qy 1025 -----DAEEYLVPOGFFCPDPAPGAGGMVHRRHSSTRSGGGDLTLGLEPSEEBAPR 1078
Db 1147 VPSTIAETDEYLPKTRPSIMLPGSA-----VEPS-DEMPK 1182
Qy 1079 S-----PLAP-----SEGAGSDVFGDGLGMGAAGKGLQSLPDPSPLORYSEDPVPLPSE 1129
Db 1183 SLRYCKDPLKPDDETGDGKEV-----GVGGIR-----LNLPLD 1216
Qy 1130 TDGVVAPLTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKD 1189
Db 1217 EDDYLMPTCOSQ-----NQS-----TPG-----YMD 1237
Qy 1190 VFAFGGAVENPEYL-----TFQGAAPQPHPPAFSPAFDNLVYWDQDPPRGAPP 1240
Db 1238 LIGVPASVDNPEYLMGSGTQATAGLAQSGMG--PHTPP-----PP 1274
Qy 1241 STFGTPTAENPE 1253
Db 1275 NTPNGMETHQHSQ 1287
RESULT 8
Q9UK79
ID Q9UK79 PRELIMINARY; PRT; 419 AA.
AC Q9UK79;
```

```
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Herstatin.
GN HER-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99415951; PubMed=10485918;
RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
RT autoinhibitor.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF177761; AAD56009.2; -.
DR InterPro; IPR00494; EGFR_L domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR SMART; SM00261; FU; 1.
SQ SEQUENCE 419 AA; 45472 MW; FEC1BE347E2D030C CRC64;

Query Match 25.9%; Score 1775.5; DB 4; Length 419;
Best Local Similarity 85.0%; Pred. No. 3.8e-126;
Matches 340; Conservative 9; Mismatches 38; Indels 13; Gaps 4;

Qy 1 MELAAALCRWGLLALLPPGAASQVCTGTDKMLRLPASPTHLDMLRHLYGCGVVOGNNL 60
Db 1 MELAAALCRWGLLALLPPGAASQVCTGTDKMLRLPASPTHLDMLRHLYGCGVVOGNNL 60
Qy 61 ELTYLPTNASLSFLQDIQEVQ--GYVLIHNRVQVPL----QRLRIVRGTFQFEDNYAL 120
Db 61 ELTYLPTNASLSFLQDIQEVQ--GYVLIHNRVQVPL----QRLRIVRGTFQFEDNYAL 114
Qy 121 AVLNDGDPPLNNTPTVTGASPGGLREQLRLSLTILKGGVLIQRNPQLCYQDTILWKDIFH 180
Db 115 AVLNDGDPPLNNTPTVTGASPGGLREQLRLSLTILKGGVLIQRNPQLCYQDTILWKDIFH 174
Qy 181 KNNQLALTLDITNRSRACHPCSPMCKSGRCWGSSEDCQSLTRTVACGAGCARCKGPLPTD 240
Db 175 KNNQLALTLDITNRSRACHPCSPMCKSGRCWGSSEDCQSLTRTVACGAGCARCKGPLPTD 234
Qy 241 CCEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGAS 300
Db 235 CCEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGAS 294
Qy 301 CVTACPNYILSTVGSGCTLVCPHNOEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVR 360
Db 295 CVTACPNYILSTVGSGCTLVCPHNOEVTAEDGTQRCCKSKPCAR----GTHSLPRPP 349
Qy 361 AVTSANIQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQ 400
Db 350 AAVPVPLRMQPG--PAHPVLVSLRPSMDLVSAFYSPLAP 387

RESULT 9
Q8R2X1
ID Q8R2X1 PRELIMINARY; PRT; 367 AA.
AC Q8R2X1;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE Hypothetical 40.2 kDa protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
```

```

RN [1]
RA SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC027080; AAH27080.1; -
KW Hypothetical protein.
SQ SEQUENCE 367 AA; 40163 MW; 0BE03395F9E101B0 CRC64;

Query Match 25.4%; Score 1739; DB 11; Length 367;
Best Local Similarity 88.0%; Pred. No. 1.8e-123;
Matches 323; Conservative 11; Mismatches 33; Indels 0; Gaps 0;

Qy 895 MALESILRRRTHQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGRLLPQPPICT 954
Db 1 MALESILRRRTHQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGRLLPQPPICT 60

Qy 955 IDVTMIMVKMWIDSECRPRELVSERFSESRWARDPQRFVJONEDLGPASPLDSTFYRSL 1014
Db 61 IDVTMIMVKMWIDSECRPRELVSERFSESRWARDPQRFVJONEDLGPSSPMDSSTFYRSL 120

Qy 1015 LEDDDMGDLVDAEYLVPQGGFFCPDPAPGAGGMVHRRSSSTRSGGGDLTLGLPSEE 1074
Db 121 LEDDDMGDLVDAEYLVPQGGFFCPDPALGNGSTAHRRHRSARSAGGELTLGLPSEE 180

Qy 1075 EAPRSLAPSGAGSDVFDGLGMAKGLQSLPHTDPSPLQRYSEDPVPLPSETDGYV 1134
Db 181 EPPRSLAPSGAGSDVFDGLGAVGTVTKGLQSLSPHLSPQLQRYSEDPVPLPSETDGYV 240

Qy 1135 APLTCSPOPEVYNQDVRPQPPSPREGPLPAARAGATLERAKTSLPCKGNVGVKDVFAFG 1194
Db 241 APLACSPQPEVYNQDVRPQPPSPREGPLPTEGPPPIRPAAGATLERPKTSLPCKGNVGVKDVFAFG 300

Qy 1195 GAVENPEYLTPOGGAAPQHPHPPAFSPAFDNLVYWDQPPRGPAPPSTFKGTPTAENPEY 1254
Db 301 GAVENPEYLA PRAGTASQPHSPAFSPAFDNLVYWDQNSSESGPPPTFTGPTAENPEY 360

Qy 1255 LGLDVPV 1261
Db 361 LGLDVPV 367

RESULT 10
Q86712 ID Q86712 PRELIMINARY; PRT; 729 AA.
AC Q86712;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Polypeptide.
GN POLYPROTEIN.
OS Avian rous-associated virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RA SEQUENCE FROM N.A.
RA MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60725.1; -
DR HSSP; P03322; 1A6S.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR004028; Retro_M.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF02813; Retro_M; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.

DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 729 AA; 80649 MW; 84D2F6914EFE1D63 CRC64;

Query Match 25.1%; Score 1720; DB 15; Length 729;
Best Local Similarity 54.8%; Pred. No. 1.4e-121;
Matches 358; Conservative 75; Mismatches 118; Indels 102; Gaps 15;

Qy 575 PQNGSVTCFGEADQCVACAHYKDPFPCVACRPSGVKPDLSYMPIMKFPDEEGACQPCPI 634
Db 141 PEETATPKTGP--DHCMMKCAHFIDGPHCVKACAGVLGENDTL-VWKYADANAVCOLCHP 197

Qy 635 NTHSCVDLLDKGCPAQASPLTISVAVV-GILLVVLGVVFGIILIKRQOKIRKYTM 693
Db 198 NCTRGCKGPGCGP---NGSKTPSTAAGVVGGLLCLVWVGLGIGLYLRRR-HIVKRTL 253

Qy 694 RRLLOETELVEPLTPSGAMNQOMRILKETELRKVKVLGSGARGTVYKGIWIPDGENVK 753
Db 254 RRLLOETELVEPLTPSGEAPNOAHLRLKETEFKVKVGLSGAPGTYYKGLWIPEGEKVK 313

Qy 754 IPVAIKVLRRENTSPKANKEILDEAYVMAGVSPVSRLLGLCLTSTVQLVTQLMPYGCLL 813
Db 314 IPVAIKELRENTSPKANKEILDEAYVMASVDNPRVCELLGLCLTSTVQLITQLMPYGCLL 373

Qy 814 DHVRENRLG3SODLLNWCQIAKGM5YLEDVRLVHRDLAARNVLKSPNHVKITDFGLA 873
Db 374 DYIREHKDNTG3SVLLNWCQIAKGMN5LEERLVHRDLAARNVLKTPQHVKITDFGLA 433

Qy 874 RLDDIDETEHAGGKVPKIMMALESILRRRPTHQSDVWSYGVTVWELMTFGAKPYDGIP 933
Db 434 KLLGADEKEHYHAGGKVPKIMMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIP 493

Qy 934 ARIPDLLEKGERLPQPICTIDVTMIMVKMWIDSECRPRELVSERFSESRWARDPQRFV 993
Db 494 ASHSSVLEKGERLPQPICTIDVTMIMVKMWIDSECRPRELVSERFSESRWARDPQRFV 553

Qy 994 VIQ-NEDLGPASPLDSTFYRSLLEDDMGDLVDAEYLVPQGGFFCPDPAPGAGGMVHHR 1052
Db 554 VIQGERMHLPSPTDSKFYRTLMEEDMEDIVDAEYLVPQGGFF----- 598

Qy 1053 HRSSTRSGGGDLTLGLPSEEEAPRSL-----APSEAGSDVFDGDLGMAKGLQSL 1107
Db 599 -NSPST-----SRTPLSLSLSATSNNSATNCID-----RNGQGH 631

Qy 1108 PTHDPSPLQRYSEDPVPLPSET--DGYVAPLTCSPQPEVYNQDVRPQPPSPREGPLPA 1165
Db 632 PVRED5FVQYISDDPTGNFLEESIDGFL-----PAPEYVQ--LMPKKPS----- 675

Qy 1166 ARPAGATLERAKTSLPCKGNVGVKDVF-----AFGGAVENPEYL 1203
Db 676 -----TAMVQNOIYNNISLTAISKLPMDSRVQNSHSTAVDNPEYL 715

RESULT 11
Q86714 ID Q86714 PRELIMINARY; PRT; 567 AA.
AC Q86714;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE V-erbB protein (Fragment).
GN V-ERBB.
OS Avian rous-associated virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RA SEQUENCE FROM N.A.
RA MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
```



```

Qy 943 KGERLPPPICTIDYIMVIMVCKWIDSECRPRFRELVSFMRMARDPQFVVIQ-NEDLG 1001
Db 354 KGERLPPPICTIDYIMVIMVCKWIDSECRPRFRELVSFMRMARDPQFVVIQDERMH 413
Qy 1002 PASPLDSTFYKSLLEDMDGLVDABEYLVPOQGFCCPDPPAGGAGMVHRRHRSSTRSG 1061
Db 414 LPSPTDSKFYRLMEEDMEDIVDAEYLVPHQGF--NPSPT--- 454
Qy 1062 GGDITLGLPSEEPSPPL-----APSEGAGSDVFDGDLGNGAAGLQSLTHDPFLQ 1116
Db 455 -----SRTPLSSLSATSNNATCIDRNGG-----H----- 481
Qy 1117 RYSDPTVPLPSETGTYVAPLTCSPQEVNQDVRPQPPSPREGPLPAARPACAT-LER 1175
Db 482 -----PVREGFL-----PAPEYVNG--LMPKKPTAMVQIQIYVLSLTAISK 523
Qy 1176 AKTILSPGKGVVYKOVFAFGGAVENPEYL 1203
Db 524 LPMDSRVQN-----SHSTAVDNPEYL 544

RESULT 15
Q9WVF5
ID Q9WVF5 PRELIMINARY; PRT; 655 AA.
AC Q9WVF5;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Epidermal growth factor receptor (Epidermal growth factor receptor
DE isoform 3).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID:10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RA Reiter J.L., Threadgill D.W., Danielson A.J., Schell C.,
RA Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
RA Maible N.J.;
RT "Alternative Transcripts from the Human and Mouse EGFR Genes Encode
RT Carboxy-Terminal Truncated Receptors.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEVTC;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Sinclair C.S., Pearshall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maible N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egr transcripts encoding truncated receptor
RT isoforms.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikola I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

```

```

RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AF124513; AAD4149.1; -.
DR EMBL; AF275366; AAG28047.1; -.
DR EMBL; AF275364; AAG28047.1; JOINED.
DR EMBL; AF275365; AAG28047.1; JOINED.
DR EMBL; AK004944; BAE23688.1; -.
DR EMBL; AK004883; BAE23641.1; -.
DR EMBL; AK004911; BAE23662.1; -.
DR MGD; MGI:95294; Egr.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR SMART; SM00261; FU; 3.
KW Receptor.
SQ SEQUENCE. 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;

Query Match 21.4%; Score 1463.5; DB 11; Length 655;
Best Local Similarity 42.7%; Pred. No. 3.3e-102;
Matches 276; Conservative 101; Mismatches 238; Indels 31; Gaps 9;

Qy 11 LLLALLPPGAA--STQVCTGTDMLRLPASPETHDMLRLHLYOGCQVQGNLEILYLP 68
Db 14 LLLTALCAAGGALBEKVCQGTSNRLTQGTGFEDHFLSLQRMNVCVVLGNLEITYV 73
Qy 69 ASLSFLQDIOEVO-FNNFTVSWLRVPKVSASHLEQRLRIVRGTLQFEDNVAVLNDG 127
Db 74 YDLSFLTKTIOEVAGYVLIANTVERPL-----ENLQIRGNALYENTYALAILSN-- 124
Qy 128 PLANTTPVTGASPGGLRELQRLSLTEILKGVLIQIRNPQLCYODTILWKDI----FHKN 183
Db 125 -----YGTNRTGLRELPMRLQELIGAVRFSNNPILCNMDTIQWRDIVQVFNMSNM 176
Qy 184 QLALTIDTNRSPACHPCSPKSGSCWGESSEDCQSLTRTVGAGCA-RCKGPLPTDCC 242
Db 177 SMDL----QSHPSPCPKDSCPNCGWGGGEECNCKLTKIIQAQCSHRCGRSPSDCC 232
Qy 243 HECCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPGRYTFGASCV 302
Db 233 HNQAAGCTGPRESDCLVCKQFODEATCKTCPLMLYNTTYQMDVNPBGKYSFGATCV 292
Qy 303 TACFYNLYSTDVGSCTLVCPHMQEVAEDGTORCEKSKPCARVCYGLGMEHLREVAV 362
Db 293 KKCPRYNVVTDHSGVRCVACGPDYVEV-EEDGIRKCKKDCGCRKVCNGIGIGEFKDTLSI 351
Qy 363 TSANIQEPACCKIFGSLAPLPESFDGDPASNTAPLOPEQLQVFTLEETLGYLYISAMP 422
Db 352 NATNIKHFKYCTAISGLHILPVAFGDSFTRTPPLDPRELEILKTKVETLGFLLIQAMP 411
Qy 423 DSLPDLISVFONLQVIRGRLHNGAYSLTLQGLGSLMGLSLRELGLSLIHHNTHLCF 482
Db 412 DNTDLHAFENLEIRGTRKHQGFSLAVGLNITSLGLSLKEISDGVIIISGNRLCY 471
Qy 483 VHTVPMDQLFRNPQALLHTANPEDECVGEGELACHOLCARGHCWGPGPTQCVNCOSFLR 542
Db 472 ANTINWKKLFGTNPQTKIMNNAEKDKAVNHVNCNPLCSSEGCWGPDEPRDCVSNVR 531
Qy 543 GQECVECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFPC 602
Db 532 GRECVKCNLTLEGPREFVENSEICQHPCLPQAMNITCTGRGPDNCIOCAHYIDGPHC 591
Qy 603 VARCPGVKPDLSYMPYTKFPEDEGACQPCPINCTHSCVDLDKGC 648
Db 592 VKTCPAGIMGENYTL-VWKYADANNVCHLCHANCYTCACGPGLOGC 636

Search completed: July 22, 2003, 09:00:11
Job time : 53.5887 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:11:38 ; Search time 38.1589 Seconds
(without alignments)
4403.399 Million cell updates/sec

Title: SEQ4-59-73-14

Perfect score: 6847

Sequence: 1 MELAALCRWGLLLALLPPG.....TFKGTPTAENPEYGLDVPV 1261

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq 101002.*

1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6713	98.0	1255	21	Human heregulin 2
2	6713	98.0	1255	22	Human heregulin 2
3	6713	98.0	1255	22	HER2 transgene pla
4	6713	98.0	1255	23	Human HER2 (ErbB2)
5	6707	98.0	1255	17	HER-2/neu protein.
6	6707	98.0	1255	20	Human HER-2/neu on
7	6707	98.0	1255	21	Human HER-2/neu pr
8	6707	98.0	1255	21	Amino acid sequenc
9	6707	98.0	1255	22	Human HER-2/neu pr
10	6707	98.0	1255	22	HER2/neu amino aci

11	6707	98.0	1255	23	AAE24067	Human Her-2 protei
12	6707	98.0	1255	23	AAE20479	Human Her-2/neu pr
13	6707	98.0	1255	23	AAE51143	Human Her-2/neu pr
14	6707	98.0	1255	23	AAU77114	Human Her-2/neu po
15	6664	97.3	1433	14	AAE39568	Sequence of c-erbB
16	6543	95.6	1223	23	AAU98923	Human breast cance
17	6390	93.3	1200	21	AAE21208	Human HER-2/neu pr
18	5911.5	86.3	1256	21	AAE21199	Rat HER-2/neu prot
19	5911.5	86.3	1256	23	AAE51144	Rat Her-2/neu onco
20	5891.5	86.0	1256	21	AAE21206	Mouse Her-2/neu pr
21	5891.5	86.0	1256	22	AAE62860	Amino acid sequenc
22	5891.5	86.0	1256	23	AAE51151	Mouse Her-2/neu on
23	4793	70.0	919	21	AAE21203	Human HER-2/neu fu
24	4793	70.0	919	23	AAE51148	Her-2/neu extracel
25	4041.5	59.0	920	23	AAE51152	Mouse Her-2/neu ex
26	4041.5	59.0	926	23	AAE51153	Mouse Her-2/neu ex
27	3677	53.7	712	21	AAE21204	Human HER-2/neu fu
28	3677	53.7	712	23	AAE51149	Her-2/neu extracel
29	3531	51.6	782	18	AAE19764	Her2-GM-CSF immuno
30	3529	51.5	653	21	AAE21200	Extracellular HER-
31	3529	51.5	653	23	AAE51145	Human Her-2/neu on
32	3491	51.0	645	23	AAE60408	Human ErbB2 oncopr
33	3491	51.0	645	22	AAE61593	Human ErbB2 extrac
34	3426	50.0	951	21	AAE44993	DC88CFV-erbB2EC fu
35	3323	48.5	624	11	AAE08222	Extracellular port
36	3098	45.2	1210	21	AAE19259	Amino acid sequenc
37	3098	45.2	1210	21	AAE50616	Human EGF receptor
38	3098	45.2	1210	23	AAE23019	Human Her-1 protei
39	3098	45.2	1210	23	AAE50768	Human epidermal gr
40	3096	45.2	1210	22	AAE68420	Amino acid sequenc
41	3084	45.0	583	23	AAE20483	Human protein for
42	3084	45.0	587	23	AAE20481	Human protein for
43	3083	45.0	589	23	AAE20484	Human protein for
44	3083	45.0	600	23	AAE20482	Human protein for
45	3057	44.6	1210	23	ABP51768	Human epidermal gr

ALIGNMENTS

RESULT 1

AAE92620
ID AAE92620 standard; Protein; 1255 AA.
XX
AC AAE92620;
XX
DT 10-AUG-2000 (first entry)
XX
DE Human heregulin 2 (Her2).
XX
KW Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;
KW self-protein; cancer; breast cancer; prostate cancer;
KW cell-associated peptide antigen; foreign epitope.
XX
OS Homo sapiens.
XX
PH Key
FT Domain
FT Location/Qualifiers
FT /label= N-terminal
FT /note= "mature polypeptide"
FT 5...25
FT Region
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 59...73
FT Region
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 103...117
FT Region
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 149...163
FT Region
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 174...323
FT Domain

FT Region /label= Cysteine_rich_domain
 FT 210..224
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 250..264
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 324..483
 FT /label= Ligand_binding_domain
 FT 325..339
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 369..383
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 465..479
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 484..623
 FT /label= Cysteine_rich_domain
 FT 579..593
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 624..654
 FT /label= Transmembrane_domain
 FT 632..652
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 653..667
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 655..1010
 FT /label= Tyrosine_kinase_domain
 FT 661..675
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 695..709
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 710..730
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 1011..1235
 FT /label= C-terminal_domain
 XX WO200020027-A2.
 XX 13-APR-2000.
 XX 05-OCT-1999; 99WO-DK00525.
 XX 05-OCT-1998; 98DK-0001261.
 XX 20-OCT-1998; 98US-0105011.
 XX (MBBI-) M & E BIOTECH AS.
 XX Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I;
 XX Gautam A, Birk P, Karlsson G;
 XX WPI; 2000-349917/30.
 XX N-PSDB; AAA09455.
 XX Inducing immune responses to weakly immunogenic, tumor associated
 XX peptide antigens for the treatment of breast and prostate cancer
 XX Claim 62; Page 193-198; 220pp; English.
 XX This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of
 XX Her2 can be used in the claimed method as an autovaccine to induce a CTL
 XX response. Subdominant CTL epitopes, antibody binding regions and
 XX cysteine residues involved in disulfide bonds are preserved in the
 XX immunogenized forms. Regions suitable for the insertion of foreign T
 XX helper epitopes were identified (see features table). The method

CC is used for inducing immune responses against weakly immunogenic
 CC cell-associated peptide antigens (PA) such as those associated with
 CC cancers (self-proteins), e.g. human prostate specific membrane antigen
 CC (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).
 CC The method comprises effecting simultaneous presentation by antigen
 CC producing cells (APCs) of the animals immune system of: (1) at least 1
 CC CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1
 CC B-cell group derived from the cell-associated PA; and (2) at least 1
 CC first T helper cell group which is foreign to the animal. Analogues of
 CC human PSM, human Her2 and human/murine FGF8b comprising a substantial
 CC part of all known and predicted CTL and B-cell epitopes of the respective
 CC PA and including at least one foreign T helper epitope are also claimed.
 CC The method is used to treat prostate, prostate/breast or breast cancer
 CC when the PA is human PSM, FGF8b and Her2, respectively.

XX Sequence 1255 AA;

Query Match 98.0%; Score 6713; DB 21; Length 1255;
 Best Local Similarity 98.5%; Pred. No. 0;
 Matches 1242; Conservative 4; Mismatches 9; Indels 6; Gaps 2;
 Qy 1 MELAALCRWGLLLALLPPGAASQVCTGDMKRLPASPETHLDMLRHLTGCCQVQGNL 60
 Db 1 MELAALCRWGLLLALLPPGAASQVCTGDMKRLPASPETHLDMLRHLTGCCQVQGNL 60
 Qy 61 ELTYLPTNASLSFLQDIQEVQFNFTVSFWLRYPKVSASHLEQRIRIVRGTLQFEDNYAL 120
 Db 61 ELTYLPTNASLSFLQDIQEVQ--GYVLIANVRQVPL----QRLRIVRGTLQFEDNYAL 114
 Qy 121 AVLNDGDPPLNNTTPVTGASPGGLRELQRLSLTELKGGVLIQRNPQLCYQDTILWKDIFH 180
 Db 115 AVLNDGDPPLNNTTPVTGASPGGLRELQRLSLTELKGGVLIQRNPQLCYQDTILWKDIFH 174
 Qy 181 KNNQALTLIDTNRSRACHPCSPMCKGSRCSWGESSEDCQSLTRTVCAAGCARCKGPLPTD 240
 Db 175 KNNQALTLIDTNRSRACHPCSPMCKGSRCSWGESSEDCQSLTRTVCAAGCARCKGPLPTD 234
 Qy 241 CCHQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGYRTFGAS 300
 Db 235 CCHQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGYRTFGAS 294
 Qy 301 CVTACPNYLSTDVGSCTLVCPLNHQNVEAEDGTQRCCKSPKPCARVCYGLGMEHLREVR 360
 Db 295 CVTACPNYLSTDVGSCTLVCPLNHQNVEAEDGTQRCCKSPKPCARVCYGLGMEHLREVR 354
 Qy 361 AVTSANTQEFAGCKKIFGSLAFLPESPDGPASNTAPLQPEQLQVFTLEBITGYLYISA 420
 Db 355 AVTSANTQEFAGCKKIFGSLAFLPESPDGPASNTAPLQPEQLQVFTLEBITGYLYISA 414
 Qy 421 WPSLPLDSLVFQNLQVIRGRILHNGAYSLLTQGLGISWLGRLSRLGSLALIHNNTHL 480
 Db 415 WPSLPLDSLVFQNLQVIRGRILHNGAYSLLTQGLGISWLGRLSRLGSLALIHNNTHL 474
 Qy 481 CFVHTVPWDQLFRNPHQALLHTANRPEDECVGEGEGLACHQLCARGHCWGPGTQCVCNSQF 540
 Db 475 CFVHTVPWDQLFRNPHQALLHTANRPEDECVGEGEGLACHQLCARGHCWGPGTQCVCNSQF 534
 Qy 541 LRQGEVCEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDDPP 600
 Db 535 LRQGEVCEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDDPP 594
 Qy 601 FCVARCPSGVKPDLSYMPIWKFDPDEGACOPCPINCTHSCVDLDDKCPAPORASPLTSI 660
 Db 595 FCVARCPSGVKPDLSYMPIWKFDPDEGACOPCPINCTHSCVDLDDKCPAPORASPLTSI 654
 Qy 661 VSAVVGILLVVVLGVWFVGLIKRRQKIRKYTMERLLQETELVEPLTPSGAMPNQAQMRI 720
 Db 655 VSAVVGILLVVVLGVWFVGLIKRRQKIRKYTMERLLQETELVEPLTPSGAMPNQAQMRI 714
 Qy 721 LKETELRKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKULRENTSPKANKEILDEAYVM 780
 Db 715 LKETELRKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKULRENTSPKANKEILDEAYVM 774

Qy 781 AGVSPVSRLLGICLTSTVOLTPQLMPYGCGLLDHVRNRRLGSDQLLNWCMQIAKMS 840
 Db 775 AGVSPVSRLLGICLTSTVOLTPQLMPYGCGLLDHVRNRRLGSDQLLNWCMQIAKMS 834
 Qy 841 YLEDRVLRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKWWALESI 900
 Db 835 YLEDRVLRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKWWALESI 894
 Qy 901 LRRRFTHQSDVMSYGVTVWELMTGAKPYDGPAREIPDLLEKGERLPQPPICITIDVYMI 960
 Db 895 LRRRFTHQSDVMSYGVTVWELMTGAKPYDGPAREIPDLLEKGERLPQPPICITIDVYMI 954
 Qy 961 MVKCMWIDSECRPRFRELVSFMSRMARDPQRFVVIQNEIDLPASPLDSTFYRSLLEDDDM 1020
 Db 955 MVKCMWIDSECRPRFRELVSFMSRMARDPQRFVVIQNEIDLPASPLDSTFYRSLLEDDDM 1014
 Qy 1021 GDLVDAEYLVPOQGFCCPDPAAGAGGVHRRHSSSTRSGGDLTLGLEPSEEAAPRSP 1080
 Db 1015 GDLVDAEYLVPOQGFCCPDPAAGAGGVHRRHSSSTRSGGDLTLGLEPSEEAAPRSP 1074
 Qy 1081 LAPSEGAGSDVFDGDLGMAAGKQLSLPTHDPSPQLQRYSEDPVPLPSETDGYVAPLTCS 1140
 Db 1075 LAPSEGAGSDVFDGDLGMAAGKQLSLPTHDPSPQLQRYSEDPVPLPSETDGYVAPLTCS 1134
 Qy 1141 POPEYVNPQDVRPOPSPREGPLPAAPAGATLERAKTLSPGKNGVVKDVPFAGGAVENP 1200
 Db 1135 POPEYVNPQDVRPOPSPREGPLPAAPAGATLERAKTLSPGKNGVVKDVPFAGGAVENP 1194
 Qy 1201 EYLTPOGGAAPQHPHFAFPAFONLYYWDODPPERGAPSTFKGTPTAENPEYGLGDVP 1260
 Db 1195 EYLTPOGGAAPQHPHFAFPAFONLYYWDODPPERGAPSTFKGTPTAENPEYGLGDVP 1254
 Qy 1261 V 1261
 Db 1255 V 1255

RESULT 2
 AAEL12130
 ID AAE12130 standard; Protein; 1255 AA.
 XX
 AC AAE12130;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Human tyrosine kinase-type receptor, HER-2.
 XX
 KW Therapeutic compound; major histocompatibility complex; vaccine;
 KW antigenic peptide; MHC; immunoregulatory; immune response; HER-2;
 KW adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;
 KW antigen presenting cell; human; tyrosine kinase-type receptor.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Region 774..782
 FT /note= "Antigenic epitope"
 XX
 WO200168677-A2.
 XX
 PD 20-SEP-2001.
 XX
 XX 16-MAR-2001; 2001WO-US40328.
 XX
 XX 16-MAR-2000; 2000US-0527487.
 XX
 XX (GENZ) GENZYME CORP.
 PA
 XX Nicolette CA;
 XX
 DR WPI: 2001-616284/71.
 DR N-PSDB; AAD19731.
 XX

PT Novel synthetic therapeutic compound for inducing immune response and
 PT for use in adoptive immunotherapy, has enhanced binding to major
 PT histocompatibility molecules and enhanced immunoregulatory properties
 XX
 PS Claim 4; Page 63-67; 69pp; English.
 XX
 CC The invention relates to synthetic therapeutic compounds (antigenic
 CC peptides) with enhanced binding to major histocompatibility complex
 CC (MHC) molecules and enhanced immunoregulatory properties relative
 CC to their natural counterparts. Compounds of the invention are useful
 CC for inducing an immune response in a subject and for use in adoptive
 CC immunotherapy. They are useful as components of anti-cancer vaccines
 CC and to expand immune effector cells that are specific for cancers
 CC characterised by expression of the breast cancer antigen, HER-2.
 CC Polynucleotides that encode peptides of the invention are useful as
 CC hybridisation probes and as primers for the detection of genes of gene
 CC transcripts that are expressed in antigen presenting cells (APCs), to
 CC confirm transduction of polynucleotides into host cells. The present
 CC sequence is human tyrosine kinase-type receptor, HER-2. Compounds
 CC of the invention are designed based on the HER-2 antigenic peptide
 CC (774-782).

XX Sequence 1255 AA;

Query Match 98.0%; Score 6713; DB 22; Length 1255;
 Best Local Similarity 98.5%; Pred. No. 0;
 Matches 1242; Conservative 4; Mismatches 9; Indels 6; Gaps 2;

Qy 1 MELAALCRWGLLALLPPGAASQVCTGDMKRLPASPETHLDMLRHLYGCGQVQGNL 60
 Db 1 MELAALCRWGLLALLPPGAASQVCTGDMKRLPASPETHLDMLRHLYGCGQVQGNL 60
 Qy 61 ELTYLPTNASLSFLQDIQEVQFNFTVSFWLRVPKVSASHLEQRLVRGTQLFEDNYAL 120
 Db 61 ELTYLPTNASLSFLQDIQEVQ--GYVLIANQVRQVPL---QKRLVRGTQLFEDNYAL 114
 Qy 121 AVLDNGDPLNNTTPVTGASFGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFH 180
 Db 115 AVLDNGDPLNNTTPVTGASFGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFH 174
 Qy 181 KNNQALTLIDINRSRACHPCSPMKGSRGWGSSSDCQSLTRTVAGGACARCKGPLPTD 240
 Db 175 KNNQALTLIDINRSRACHPCSPMKGSRGWGSSSDCQSLTRTVAGGACARCKGPLPTD 234
 Qy 241 CCHQCAAGCT3PKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGAS 300
 Db 235 CCHQCAAGCT3PKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGAS 294
 Qy 301 CVTACPNYLSLTDVGSCTLVCPLNHNOEVTAEADGTQRCCKSPKPCARVCYGLGMEHLREVR 360
 Db 295 CVTACPNYLSLTDVGSCTLVCPLNHNOEVTAEADGTQRCCKSPKPCARVCYGLGMEHLREVR 354
 Qy 361 AVTSANIQEPAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFTLEITGYLYISA 420
 Db 355 AVTSANIQEPAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFTLEITGYLYISA 414
 Qy 421 WPDLSPLSVFQNLQVIRGRILHNGAYSLTLQGLISWLGLRSLRELGSGLALHNNHNL 480
 Db 415 WPDLSPLSVFQNLQVIRGRILHNGAYSLTLQGLISWLGLRSLRELGSGLALHNNHNL 474
 Qy 481 CFVHTVPWDQLFRNPHOALLHTANRPEDECVGEGLAGCHQLCARGHCWGPGPTQCVNCSQF 540
 Db 475 CFVHTVPWDQLFRNPHOALLHTANRPEDECVGEGLAGCHQLCARGHCWGPGPTQCVNCSQF 534
 Qy 541 LRGOECVEECRLVQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDDP 600
 Db 535 LRGOECVEECRLVQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDDP 594
 Qy 601 FCVARCPGSKVDLSYMPIWKFPDEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTSI 660
 Db 595 FCVARCPGSKVDLSYMPIWKFPDEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTSI 654

```
Qy 661 VSAVVGILLVVVGWFGILLKRRQOKIRKYTMRLLOQETELVPLTPSGAMPNQAQMRI 720
Db |||||
Qy 655 VSAVVGILLVVVGWFGILLKRRQOKIRKYTMRLLOQETELVPLTPSGAMPNQAQMRI 714
Db |||||
Qy 721 LKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVM 780
Db |||||
Qy 715 LKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVM 774
Db |||||
Qy 781 AGVSPVYSRLIGLCSTVQLVLTQMLPYGCLLDHVRNRRGLSQDILLNMCQIAKMS 840
Db |||||
Qy 775 AGVSPVYSRLIGLCSTVQLVLTQMLPYGCLLDHVRNRRGLSQDILLNMCQIAKMS 834
Db |||||
Qy 841 YLEVDRLVHRDLAARNVLKSPNVKVTDFGLARLLDIDETEHADGGKVPKMALES1 900
Db |||||
Qy 835 YLEVDRLVHRDLAARNVLKSPNVKVTDFGLARLLDIDETEHADGGKVPKMALES1 894
Db |||||
Qy 901 LRRRFTHQSDVMSYGVTVWELMTFCAKPYDGIIPAREIPDLEKGERLPQPPICTIDVYMI 960
Db |||||
Qy 895 LRRRFTHQSDVMSYGVTVWELMTFCAKPYDGIIPAREIPDLEKGERLPQPPICTIDVYMI 954
Db |||||
Qy 961 MVKCMWIDSECRPRFRELVSFSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDM 1020
Db |||||
Qy 955 MVKCMWIDSECRPRFRELVSFSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDM 1014
Db |||||
Qy 1021 GDLVDAEYLVPQGFPCPDPAAGAGGMVHRHSSSTRSGGDLTLGLEPSEEEAPRSP 1080
Db |||||
Qy 1015 GDLVDAEYLVPQGFPCPDPAAGAGGMVHRHSSSTRSGGDLTLGLEPSEEEAPRSP 1074
Db |||||
Qy 1081 LAPSEGAGSDVFDGDLGMAAKGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCS 1140
Db |||||
Qy 1075 LAPSEGAGSDVFDGDLGMAAKGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCS 1134
Db |||||
Qy 1141 POPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTLPSPKNGVVKDVPFAGGAVENP 1200
Db |||||
Qy 1135 POPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTLPSPKNGVVKDVPFAGGAVENP 1194
Db |||||
Qy 1201 EYLTPQGGAAPOHPHPPAFSFAFNLYYWDQDPPERGAPPSTFKGTPPTAENPEYLGLDVP 1260
Db |||||
Qy 1195 EYLTPQGGAAPOHPHPPAFSFAFNLYYWDQDPPERGAPPSTFKGTPPTAENPEYLGLDVP 1254
Db |||||
Qy 1261 V 1261
Db 1255 V 1255

RESULT 3
AAB60167
ID AAB60167 standard; Protein; 1255 AA.
AC
XX
AC AAB60167;
XX
DT 03-APR-2001 (first entry)
XX
DE HER2 transgene plasmid construct encoded protein.
XX
KW Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
KW antibody.
XX
OS Homo sapiens.
OS Synthetic.
XX
XX WO200100244-A2.
XX
XX 04-JAN-2001.
XX
XX 23-JUN-2000; 2000WO-US17229.
XX
XX 25-JUN-1999; 99US-0141316.
XX
XX 16-MAR-2000; 2000US-0189844.
XX
XX (GETH ) GENENTECH INC.
XX
XX Erickson,S, Schwall R;
```

```
XX WPI; 2001-061962/07.
DR N-PSDB; AAF24297.
XX
PT Treating tumors, particularly breast cancers, which overexpress an ErbB
PT receptor and does not respond to an anti-ErbB antibody, comprises
PT conjugating the antibody to a maytansinoid -
XX
PS Example 3; Fig 4; 92pp; English.
XX
CC The present invention provides a method of treating cancer by
CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In
CC particular, the antibody is directed against ErbB2 (also known as HER2
CC and p185neu). The method is particularly useful in the treatment of
CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.
XX
SQ Sequence 1255 AA;
Query Match 98.0%; Score 6713; DB 22; Length 1255;
Best Local Similarity 98.5%; Pred. No. 0;
Matches 1242; Conservative 4; Mismatches 9; Indels 6; Gaps 2;
Qy 1 MELAALCRWGLLLALLPPGAASQVCTGTDKRLRPASPETHDMLRHLYGQCVVQGNL 60
Db |||||
Qy 1 MELAALCRWGLLLALLPPGAASQVCTGTDKRLRPASPETHDMLRHLYGQCVVQGNL 60
Db |||||
Qy 61 ELYVLPNASLSFLQDIQEVQFNFTVSVFVRLPKVSASHLEQRLRIVRGTLQEDNYAL 120
Db |||||
Qy 61 ELYVLPNASLSFLQDIQEVQ--GYVLIHNRVQVPL---QRLRIVRGTLQEDNYAL 114
Db |||||
Qy 121 AVLDNGDPLNNTTPTVTCASPGGLRELQRLSITELTKGGVLIQRPOLCYQDTILWKDIFH 180
Db |||||
Qy 115 AVLDNGDPLNNTTPTVTCASPGGLRELQRLSITELTKGGVLIQRPOLCYQDTILWKDIFH 174
Db |||||
Qy 181 KNNQALTLTDNRSRACHPCSPMKGSRGWSESSDCSLTRTVTCAGGCARCKGPLPTD 240
Db |||||
Qy 175 KNNQALTLTDNRSRACHPCSPMKGSRGWSESSDCSLTRTVTCAGGCARCKGPLPTD 234
Db |||||
Qy 241 CCHQCAAGCTGPKHSDCLACLPNHSIGICELHCPALVTYNTDTFESMPNPEGYTFGAS 300
Db |||||
Qy 235 CCHQCAAGCTGPKHSDCLACLPNHSIGICELHCPALVTYNTDTFESMPNPEGYTFGAS 294
Db |||||
Qy 301 CVTACPNYLSLTDVGSCTLVCPLNHQNVTADGTQRCCKSPKPCARVCYGLGMHLEVR 360
Db |||||
Qy 295 CVTACPNYLSLTDVGSCTLVCPLNHQNVTADGTQRCCKSPKPCARVCYGLGMHLEVR 354
Db |||||
Qy 361 AVTSANIQEFAGCKKIFGSLAFLPESPDGDPASNTAPLQPEQLQVFTLEETITGYLYISA 420
Db |||||
Qy 355 AVTSANIQEFAGCKKIFGSLAFLPESPDGDPASNTAPLQPEQLQVFTLEETITGYLYISA 414
Db |||||
Qy 421 WPDSPDLVSFQNLQVIRGRILHNGAYSLTLQGLIGISWLGRLSRLRELSGLALITHNTHL 480
Db |||||
Qy 415 WPDSPDLVSFQNLQVIRGRILHNGAYSLTLQGLIGISWLGRLSRLRELSGLALITHNTHL 474
Db |||||
Qy 481 CFVHTVPWDOLFNRPHQALLHTANRDECVGEGCLACHOLCARGHCHGPGTQCVCNCSQF 540
Db |||||
Qy 475 CFVHTVPWDOLFNRPHQALLHTANRDECVGEGCLACHOLCARGHCHGPGTQCVCNCSQF 534
Db |||||
Qy 541 LRQECVVEECRVLQGLPREYVYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDDP 600
Db |||||
Qy 535 LRQECVVEECRVLQGLPREYVYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDDP 594
Db |||||
Qy 601 FCVARCPSGVKPDLSYMPYIWKFPDEEGACQPCPINCSTHSCVDLDDKDGCPAQRASPLTSI 660
Db |||||
Qy 595 FCVARCPSGVKPDLSYMPYIWKFPDEEGACQPCPINCSTHSCVDLDDKDGCPAQRASPLTSI 654
Db |||||
Qy 661 VSAVVGILLVVVGWFGILLKRRQOKIRKYTMRLLOQETELVPLTPSGAMPNQAQMRI 720
Db |||||
Qy 655 VSAVVGILLVVVGWFGILLKRRQOKIRKYTMRLLOQETELVPLTPSGAMPNQAQMRI 714
Db |||||
Qy 721 LKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVM 780
Db |||||
```

Db 715 LKETELRKVKVLGSGAFQTVVKGWIPDGENVKIPVAIKVIRENTSPKANKIILDEAYVM 774
Qy 781 AGVGSPPVSRLLGICLTSTVQLMPYGLLDHVRNRRGLASQDILLNWCMIAGMS 840
Db 775 AGVGSPPVSRLLGICLTSTVQLMPYGLLDHVRNRRGLASQDILLNWCMIAGMS 834
Qy 841 YLEDVRLVHRLAARNVLKSPNVHKITDFGLARLLDIDETEHADGGKVPKMALES 900
Db 835 YLEDVRLVHRLAARNVLKSPNVHKITDFGLARLLDIDETEHADGGKVPKMALES 894
Qy 901 LRRRFTHQSDVWSYGVTVWELMTGAKPYDGI PAREIPDLLEKGERLPQPPICITDVYMI 960
Db 895 LRRRFTHQSDVWSYGVTVWELMTGAKPYDGI PAREIPDLLEKGERLPQPPICITDVYMI 954
Qy 961 MVKCMWIDSECRPRFRELVSFSEMRARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDM 1020
Db 955 MVKCMWIDSECRPRFRELVSFSEMRARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDM 1014
Qy 1021 GDLVDAEYLYVPOQGFPCPDAPGAGGMVHRRSSSTRSGGDLITGLPSEBEAPRSP 1080
Db 1015 GDLVDAEYLYVPOQGFPCPDAPGAGGMVHRRSSSTRSGGDLITGLPSEBEAPRSP 1074
Qy 1081 LAPSEGAGSDVFDGDLGMAAGKQSLPHTDPSPLQRYSEDPVPLPSETDGYVAPLTC 1140
Db 1075 LAPSEGAGSDVFDGDLGMAAGKQSLPHTDPSPLQRYSEDPVPLPSETDGYVAPLTC 1134
Qy 1141 PQPEYVNPQDVRPQPPREGPLPAARPAGATLBRAKTLPSPKNGVWVDVFAFGAVENP 1200
Db 1135 PQPEYVNPQDVRPQPPREGPLPAARPAGATLBRAKTLPSPKNGVWVDVFAFGAVENP 1194
Qy 1201 EYLTPQGAAPQPHPPAPSPAFONLYYWDQDPPERGAPPTFKGTPTAENPEYGLDVP 1260
Db 1195 EYLTPQGAAPQPHPPAPSPAFONLYYWDQDPPERGAPPTFKGTPTAENPEYGLDVP 1254
Qy 1261 V 1261
Db 1255 V 1255

RESULT 4
AAU74545
ID AAU74545 standard; Protein; 1255 AA.
XX
AC AAU74545;
XX
DT 23-APR-2002 (first entry)
XX
DE Human HER2 (Erbb2) polypeptide.
XX
KW Human; HER2; Erbb; epidermal growth factor receptor; receptor;
KW anti-Erbb antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
KW thyroid; pancreas; prostate; bladder; Erbb2; neuronal disorder;
KW glial disorder; astrocytal disorder; hypothalamic disorder;
KW glandular disorder; macrophagal disorder; epithelial disorder;
KW stromal disorder; blastocoealic disorder; inflammatory disorder;
KW angiogenic disorder; immunological disorder.
XX
OS Homo sapiens.
XX
PN US2002001587-A1.
XX
PD 03-JAN-2002.
XX
PF 16-MAR-2001; 2001US-0811123.
XX
PR 16-MAR-2000; 2000US-189844P.
PR 05-OCT-2000; 2000US-238327P.
XX
PA (ERIC/) BRICKSON S.
PA (SCHW/) SCHWALL R.
PA (SLIW/) SLIWKOWSKI M.
XX

PI Erickson S, Schwall R, Sliwkowski M;
XX WPI; 2002-163686/21.
DR N-PSDB; ABK14058.
XX
PT Treating tumour characterised by overexpression of epidermal growth
PT factor receptor, Erbb or cancer in mammal, comprises administering
PT anti-Erbb antibody-maytansinoid conjugate to the mammal -
XX
PS Example 3; Fig 7; 93pp; English.
XX
CC The invention relates to treating a tumour in a mammal, where the tumour
CC is characterised by the overexpression of an epidermal growth factor
CC receptor (Erbb) and does not respond or responds poorly, to treatment
CC with an anti-Erbb antibody, comprising administering to the mammal an
CC anti-Erbb antibody-maytansinoid conjugate. The method is useful for
CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
CC prostate and bladder, preferably breast cancer. The breast cancer is a
CC metastatic breast cancer or an aggressive form of metastatic breast
CC cancer which overexpresses Erbb2. The method is also useful for treating
CC neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal,
CC epithelial, stromal, blastocoealic, inflammatory, angiogenic and
CC immunological disorders. This sequence represents the human HER2 (Erbb2)
CC polypeptide of the invention.
XX
SQ Sequence 1255 AA;
Query Match 98.0%; Score 6713; DB 23; Length 1255;
Best Local Similarity 98.5%; Pred. No. 0;
Matches 1242; Conservative 4; Mismatches 9; Indels 6; Gaps 2;
Qy 1 MELAALCRWGLLALLPPGAASQVCTGTDKMLRASPETHLDMLRHLQGCGVOVGNL 60
Db 1 MELAALCRWGLLALLPPGAASQVCTGTDKMLRASPETHLDMLRHLQGCGVOVGNL 60
Qy 61 ELTYLPTNASLSFLQDIOEQVFNFTVSPWLRVPKVSASHLEQRLEIRVGTQOLFEDNYAL 120
Db 61 ELTYLPTNASLSFLQDIOEQVFNFTVSPWLRVPKVSASHLEQRLEIRVGTQOLFEDNYAL 114
Qy 121 AVLDNGDPLNNTPTVTGASPGGLRELQRLSLTILKGGVLIQRNPQLCYODTILWKDIFH 180
Db 115 AVLDNGDPLNNTPTVTGASPGGLRELQRLSLTILKGGVLIQRNPQLCYODTILWKDIFH 174
Qy 181 KKNQLALTLDITNRSRACHPCSPMGSRGSGESSEDQSLTRTVCAAGCARCKGPIPTD 240
Db 175 KKNQLALTLDITNRSRACHPCSPMGSRGSGESSEDQSLTRTVCAAGCARCKGPIPTD 234
Qy 241 CCEQCAAGCT3PKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGAS 300
Db 235 CCEQCAAGCT3PKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGAS 294
Qy 301 CVTACPNYLSLTDVGSCTLVCPHNOEVTAEDGTQRCCKSPCARVCYGLGMEHLREVR 360
Db 295 CVTACPNYLSLTDVGSCTLVCPHNOEVTAEDGTQRCCKSPCARVCYGLGMEHLREVR 354
Qy 361 AVTSANIOEPAGCKKIFGSLAFLPESDGPASNTAPLOPEQLQVFFETLEITGLYLISA 420
Db 355 AVTSANIOEPAGCKKIFGSLAFLPESDGPASNTAPLOPEQLQVFFETLEITGLYLISA 414
Qy 421 WPDLSPLDSVFNQLQVIRGRILHNGAYSLTLOGIGISWLGRLSRLRELSGLALIHNNTHL 480
Db 415 WPDLSPLDSVFNQLQVIRGRILHNGAYSLTLOGIGISWLGRLSRLRELSGLALIHNNTHL 474
Qy 481 CFVHTVPWDQLFRNPQALLHTANRBEDECVGSLACHQLCARGHCWGPGPTQCVNCSQF 540
Db 475 CFVHTVPWDQLFRNPQALLHTANRBEDECVGSLACHQLCARGHCWGPGPTQCVNCSQF 534
Qy 541 LRGOECVBECEVLOGLPREYVNHARHCLPCHPECPONGSVTCFPGPADQCVACAHYKDP 600
Db 535 LRGOECVBECEVLOGLPREYVNHARHCLPCHPECPONGSVTCFPGPADQCVACAHYKDP 594
Qy 601 FCVARCPGSKVFDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTSI 660

```
Db 595 FCVACPSGVKFDLSYMPDKWPFDEEGACQPCINCHSCVDLDDKGCPCAPQASPLTSI 654
Qy 661 VSAVVGILLVVVLGVVFGILIKRQOKIRKYMTRLLQSTELVEPLTPSGAMPNQAQMRI 720
Db 655 VSAVVGILLVVVLGVVFGILIKRQOKIRKYMTRLLQSTELVEPLTPSGAMPNQAQMRI 714
Qy 721 LKETELRKVKVLGSGAGFTVVGKTIWIPDGENVKIPVAIKVRENTSPKANKEIIDEAYVM 780
Db 715 LKETELRKVKVLGSGAGFTVVGKTIWIPDGENVKIPVAIKVRENTSPKANKEIIDEAYVM 774
Qy 781 AGVGSPPVSRLLGICLTSTVOLVTQLMPYGCCLLDHVRNRRGLSGQDILLNWCMIAGKMS 840
Db 775 AGVGSPPVSRLLGICLTSTVOLVTQLMPYGCCLLDHVRNRRGLSGQDILLNWCMIAGKMS 834
Qy 841 YLEDVRLVHRDLAARNVLKSPNHNKTIIDFGIARLLDIDETEHADGKGVPIKWMALESI 900
Db 835 YLEDVRLVHRDLAARNVLKSPNHNKTIIDFGIARLLDIDETEHADGKGVPIKWMALESI 894
Qy 901 LRRRFTHQSDVWSYGVTVWELMTTCAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYMI 960
Db 895 LRRRFTHQSDVWSYGVTVWELMTTCAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYMI 954
Qy 961 MVKCMWIDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDM 1020
Db 955 MVKCMWIDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDM 1014
Qy 1021 GDLVDAEYLVPOQGFPCDPAPGAGGMVHRRHSSSTRSGGDLITLGLEPSEEAAPRSP 1080
Db 1015 GDLVDAEYLVPOQGFPCDPAPGAGGMVHRRHSSSTRSGGDLITLGLEPSEEAAPRSP 1074
Qy 1081 LAPSEGAGSDVFDGDLGMAAKGLQSLPTHDPSPLOQRYSEDPVPLPSETDGYVAPLTCS 1140
Db 1075 LAPSEGAGSDVFDGDLGMAAKGLQSLPTHDPSPLOQRYSEDPVPLPSETDGYVAPLTCS 1134
Qy 1141 POPEYVNPQDVRPOPSPREGPLPAARPAAGATLERAKTSLSPGKNGVVKDVFAGGAVENP 1200
Db 1135 POPEYVNPQDVRPOPSPREGPLPAARPAAGATLERAKTSLSPGKNGVVKDVFAGGAVENP 1194
Qy 1201 EYLTPQGAAPQHPHPPAFSFDNLVYWDODPPERGAPPTFKGTPTAENPEYLGLDVP 1260
Db 1195 EYLTPQGAAPQHPHPPAFSFDNLVYWDODPPERGAPPTFKGTPTAENPEYLGLDVP 1254
Qy 1261 V 1261
Db 1255 V 1255

RESULT 5
AAW01111
ID AAW01111 standard; Protein; 1255 AA.
XX AC AAW01111;
XX AC AAW01111;
DT 01-JAN-1997 (first entry)
XX DE HER-2/neu protein.
XX KW HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
XX KW breast cancer; ovary cancer; colon cancer; lung cancer;
XX KW prostate cancer; immunisation; tumour; vaccine; vector.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT Domain 676..1255
XX FT /label= Intracellular domain
XX FT /note= "claimed domain, useful for immunisation"
XX PN W09630514-A1.
XX PD 03-OCT-1996.
XX
```

```
PF 28-MAR-1996; 96WO-US01689.
XX 31-MAR-1995; 95US-0414417.
XX (UNIW ) UNIV WASHINGTON.
XX Cheever MA, Disis ML;
XX MPI; 1996-455361/45.
XX N-PSDB; AAT40739.
XX DNA encoding HER-2-neu poly:peptide(s) - used for prevention or
XX treatment of malignancies with which the HER-2/neu oncogene is
XX associated
XX Claim 2; Page 56-61; 71pp; English.
XX Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is
XX the product of the HER-2/neu oncogene (see also AAT40739). The
XX protein is over-expressed in various cancers, including breast,
XX ovarian, colon, lung and prostate. The intracellular domain of the
XX protein can be used to immunise an animal against a malignancy with
XX which the oncogene is associated. The polypeptide can be produced
XX in transformed host cells for use in immunisation. Alternatively,
XX animal cells are transfected in vivo or ex vivo with a viral vector
XX that directs expression of the polypeptide.
XX SQ Sequence 1255 AA;
```

Query Match 98.0%; Score 6707; DB 17; Length 1255;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1240; Conservative 5; Mismatches 10; Indels 6; Gaps 2;

```
Qy 1 MELAALCRWGLLALLPPGAASQVCTGDMKRLPASPETHDMLRHLVGGCQVVOGNL 60
Db 1 MELAALCRWGLLALLPPGAASQVCTGDMKRLPASPETHDMLRHLVGGCQVVOGNL 60
Qy 61 ELTYLPNASLSFLQDIOEQVFNFTVSFWRLPKVSASHLEORLIRVRGTQLPEDNYAL 120
Db 61 ELTYLPNASLSFLQDIOEQV--GYVLIANQVRQVPL----QELRIRVGTQLPEDNYAL 114
Qy 121 AVLNDGDLNNTTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFH 180
Db 115 AVLNDGDLNNTTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFH 174
Qy 181 KNNQALTLIDTNRSRACHPCSPMCKSGRCWGSSEDCQSLTRTVACGGCARCKGPLPTD 240
Db 175 KNNQALTLIDTNRSRACHPCSPMCKSGRCWGSSEDCQSLTRTVACGGCARCKGPLPTD 234
Qy 241 CCHEOCAAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGAS 300
Db 235 CCHEOCAAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGAS 294
Qy 301 CVTACPNYLSTDVGSCTLVCPLNHNEVTAEDGTQRCCKSKPCARVCYGLGMHLEVR 360
Db 295 CVTACPNYLSTDVGSCTLVCPLNHNEVTAEDGTQRCCKSKPCARVCYGLGMHLEVR 354
Qy 361 AVTSANTQEFAGCKKIFGSLAFIPESPDGPASNTAPLOPELOVFTLEITCYLYISA 420
Db 355 AVTSANTQEFAGCKKIFGSLAFIPESPDGPASNTAPLOPELOVFTLEITCYLYISA 414
Qy 421 WPDSDLPSLVFQNLQVIRGRILHNGAYSLSLTQGLGISWLGRLSRLRELSGLALHNNTHL 480
Db 415 WPDSDLPSLVFQNLQVIRGRILHNGAYSLSLTQGLGISWLGRLSRLRELSGLALHNNTHL 474
Qy 481 CFVHTVPWDQLFRNPHOALLHTANRPEDECVGEGLAHQLCARGHCWGPGPTQCVCNSQF 540
Db 475 CFVHTVPWDQLFRNPHOALLHTANRPEDECVGEGLAHQLCARGHCWGPGPTQCVCNSQF 534
Qy 541 LRQGECEVEECRVLOGLPREYVNAHCLPCHPECOPOQNGSVTCFGEPAEQCVACAHYKDDP 600
Db 535 LRQGECEVEECRVLOGLPREYVNAHCLPCHPECOPOQNGSVTCFGEPAEQCVACAHYKDDP 594
```

```

QY 601 FCVARCPSGVKPDLSYMPDIWKFPPDEBGAQCQPCINCTHSCVDLDDKGCAPAQORASPLTSI 660
Db 595 FCVARCPSGVKPDLSYMPDIWKFPPDEBGAQCQPCINCTHSCVDLDDKGCAPAQORASPLTSI 654
QY 661 VSAVVGILLVVLGVVFGILLIKRQOKIRKTYTMRLLQETELVEPLTPSGAMPNOAQMRI 720
Db 655 ISAVVGILLVVLGVVFGILLIKRQOKIRKTYTMRLLQETELVEPLTPSGAMPNOAQMRI 714
QY 721 LKETELRKVKVLSGAGFGTVYKGIWIPDGENVKIIPVAIKVLRENTSPKANKEILLDEAYVM 780
Db 715 LKETELRKVKVLSGAGFGTVYKGIWIPDGENVKIIPVAIKVLRENTSPKANKEILLDEAYVM 774
QY 781 AGVSPYVSRLLGICLSTVQLMPLPYGCLLDHVRNRRGLSQDILLNCMOIAKMS 840
Db 775 AGVSPYVSRLLGICLSTVQLMPLPYGCLLDHVRNRRGLSQDILLNCMOIAKMS 834
QY 841 YLEDRVLRDLAARNVLKSPNVKITDFGLARLLDIDETEHADGGKVPKMALESI 900
Db 835 YLEDRVLRDLAARNVLKSPNVKITDFGLARLLDIDETEHADGGKVPKMALESI 894
QY 901 LRRRFTHQSDVMSYGVTVWELMTFCAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMI 960
Db 895 LRRRFTHQSDVMSYGVTVWELMTFCAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMI 954
QY 961 MYKCMWIDSECRPRFRELVSFSEMRARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDM 1020
Db 955 MYKCMWIDSECRPRFRELVSFSEMRARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDM 1014
QY 1021 GDLVDAEYLVPOQGFCCPDPAAGAGGMVHRRSSSTRSGGDLTLGLEPSESEAPRSP 1080
Db 1015 GDLVDAEYLVPOQGFCCPDPAAGAGGMVHRRSSSTRSGGDLTLGLEPSESEAPRSP 1074
QY 1081 LAPSEGAGSDVFDGDLGMAAGKGLQSLTPHDPSPLQRYSEDPVPLPSETDGYVAPLTC 1140
Db 1075 LAPSEGAGSDVFDGDLGMAAGKGLQSLTPHDPSPLQRYSEDPVPLPSETDGYVAPLTC 1134
QY 1141 PQPEYVNPQDVRPQPPSPREGPLPAAPAGATLERAKTLSPGKNGVVKDVPFAFGAVENP 1200
Db 1135 PQPEYVNPQDVRPQPPSPREGPLPAAPAGATLERAKTLSPGKNGVVKDVPFAFGAVENP 1194
QY 1201 EYLTPOGGAAPQHPHPPAFSPAFNLVYWDODPPERGAPPSTFKGTPTAENPEYGLDVP 1260
Db 1195 EYLTPOGGAAPQHPHPPAFSPAFNLVYWDODPPERGAPPSTFKGTPTAENPEYGLDVP 1254
QY 1261 V 1261
Db 1255 V 1255

RESULT 6
AAW92406
ID AAW92406 standard; Protein; 1255 AA.
XX AC AAW92406;
XX AC
XX DT 21-APR-1999 (first entry)
XX DE
XX DE Human HER-2/neu oncogene protein.
XX KW HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;
XX KW malignancy; treatment; tumour.
XX OS Homo sapiens.
XX OS
XX FH Location/Qualifiers
XX FT Region
XX FT 676..1255
XX FT /note= "region which elicits immune response"
XX XX
XX PN US5869445-A.
XX XX
XX PD 09-FEB-1999.
XX XX
XX PF 01-APR-1996; 96US-0625101.

```

```

XX 01-APR-1996; 96US-0625101.
PR 17-MAR-1993; 93US-0033644.
PR 12-AUG-1993; 93US-0106112.
PR 31-MAR-1995; 95US-0414417.
XX (UNIW ) UNIV WASHINGTON.
XX Cheever MA, Disis ML;
XX MPI; 1999-152835/13.
XX N-PSDB; AAX01912.
XX
PT Use of HER-2/neu polypeptides - for eliciting an immune response to
PT an HER-2/neu associated malignancy, particularly for treating or
PT preventing tumours
XX
PS Claim 3; Column 31-38; 26pp; English.
XX
CC This sequence represents the human HER-2/neu oncogene protein. A fragment
CC of this protein is used in a method for eliciting or enhancing an immune
CC response to HER-2/neu protein. The polypeptide can stimulate T cells and
CC B cells to produce an immune response to the HER-2/neu protein. The
CC method can be used for immunisation against a malignancy in which the
CC HER-2/neu oncogene is associated and in the treatment of an existing
CC tumour, or to prevent tumour occurrence or reoccurrence.
XX
SQ Sequence 1255 AA;
Query Match 98.0%; Score 6707; DB 20; Length 1255;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1240; Conservative 5; Mismatches 10; Indels 6; Gaps 2;
QY 1 MEALACRCWGLLALLPPGAASQVCTGDMKRLPASPTHLDMLRHLYQGCOVQGNL 60
Db 1 MEALACRCWGLLALLPPGAASQVCTGDMKRLPASPTHLDMLRHLYQGCOVQGNL 60
QY 61 ELTYLPTNASLSFLQDIQEVQFNFTVSFWLRVPKVSASHLEQRRLVRGTQLPEDNYAL 120
Db 61 ELTYLPTNASLSFLQDIQEVQ--GYVLIHNRVQVPL---QRLRLVRGTQLPEDNYAL 114
QY 121 AVLDNGDPLNNTTPTVGTASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFH 180
Db 115 AVLDNGDPLNNTTPTVGTASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFH 174
QY 181 KNNQLALTLDITNRSRACHPCSPMKSGSRGCSSESDCQSLTRTVCGAGCARCKGPLPTD 240
Db 175 KNNQLALTLDITNRSRACHPCSPMKSGSRGCSSESDCQSLTRTVCGAGCARCKGPLPTD 234
QY 241 CHEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGAS 300
Db 235 CHEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGAS 294
QY 301 CVTACPNYLSLTDVSGCTLVCPLNHNEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVR 360
Db 295 CVTACPNYLSLTDVSGCTLVCPLNHNEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVR 354
QY 361 AVTSANIQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFETLEITGLYLISA 420
Db 355 AVTSANIQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFETLEITGLYLISA 414
QY 421 WPDLSLPDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLRELGSGLALIHNTHL 480
Db 415 WPDLSLPDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLRELGSGLALIHNTHL 474
QY 481 CFVHTVPWDLFRNPQALLHTANRPDEBCVGBGLACHQLCARGHGCMGPGTQCVCNSQF 540
Db 475 CFVHTVPWDLFRNPQALLHTANRPDEBCVGBGLACHQLCARGHGCMGPGTQCVCNSQF 534
QY 541 LRQGECEVECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFPGPADQCACAHYKDDPP 600
Db 535 LRQGECEVECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFPGPADQCACAHYKDDPP 594

```


Db 595 FCVARGSGVDPDLSYMPIWKFPEDEGACQPCPINCTHSCVDLDDKGPASQASPLTISI 654
Qy 661 VSADVGLLVVVLGVVFGILIKRQOKIRKTYTMRLLQETELVEPLTPSGAMPNQAOWRI 720
Db 655 ISADVGLLVVVLGVVFGILIKRQOKIRKTYTMRLLQETELVEPLTPSGAMPNQAOWRI 714
Qy 721 LKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEIILDEAYVM 780
Db 715 LKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEIILDEAYVM 774
Qy 781 AGVGSPPVSRLLGICLTSTVOLTPQMPYGCCLLDHVRNRCRLGSODLLNWCMIAGMS 840
Db 775 AGVGSPPVSRLLGICLTSTVOLTPQMPYGCCLLDHVRNRCRLGSODLLNWCMIAGMS 834
Qy 841 YLEDRVLRDLAARNVLKSPNVHKITDFGLARLLDIDETEHADGGKVPDKWMALESI 900
Db 835 YLEDRVLRDLAARNVLKSPNVHKITDFGLARLLDIDETEHADGGKVPDKWMALESI 894
Qy 901 LRRRFTHQSDVMSYGVTVWELMTGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYMI 960
Db 895 LRRRFTHQSDVMSYGVTVWELMTGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYMI 954
Qy 961 MVKCMWIDSECRPRELVSFSSWARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDM 1020
Db 955 MVKCMWIDSECRPRELVSFSSWARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDM 1014
Qy 1021 GDLVDAEYLVPOQGFCDPAPGAGGMVHRHRSSTRSGGDLTLGLEPSEBEAPRSP 1080
Db 1015 GDLVDAEYLVPOQGFCDPAPGAGGMVHRHRSSTRSGGDLTLGLEPSEBEAPRSP 1074
Qy 1081 LAPSEGAGSVDFDGLGMAAKGLQSLTPHDPSPLORYSEDPTVPLPSETDGYVAPLTCS 1140
Db 1075 LAPSEGAGSVDFDGLGMAAKGLQSLTPHDPSPLORYSEDPTVPLPSETDGYVAPLTCS 1134
Qy 1141 POEYVNPQDVRPOPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAGGAVENP 1200
Db 1135 POEYVNPQDVRPOPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAGGAVENP 1194
Qy 1201 EYLTPOGGAAPQHPPPAFSPAFNLVYWDQDPPERGAAPPSTFKGTPPTAENPEYLGLDVP 1260
Db 1195 EYLTPOGGAAPQHPPPAFSPAFNLVYWDQDPPERGAAPPSTFKGTPPTAENPEYLGLDVP 1254
Qy 1261 V 1261
Db 1255 V 1255

RESULT 8
ID AAY84780 standard; Protein; 1255 AA.

AC AAY84780;
XX AC
XX AC
XX AC
DT 08-AUG-2000 (first entry)
XX
XX Amino acid sequence of the SPLICE erbb-2 receptor protein.
DE
XX SPLICE erbb-2 receptor protein; cell transformation disorder; cancer;
XX tumor cell proliferation; tissue degeneration; arthropathy;
KW bone resorption; inflammatory disease; degenerative disorder;
KW wound healing.
XX
XX Homo sapiens.
XX
XX WO200020579-A1.
XX
XX 13-APR-2000.
XX
XX 01-OCT-1999; 99WO-CA00912.
XX
XX 02-OCT-1998; 98US-0165192.
XX

PA (UYMC-) UNIV MCMASTER.
XX Muller WJ, Siegel PM;
XX WPI: 2000-303768/26.
DR N-PSDB; AAA14812.
XX
PT Nucleic acid encoding an erbb 2 receptor protein designated SPLICE
XX erbb-2, inhibitors of the protein are useful for treatment of cancer -
XX
PS Claim 3; Fig 2; 60pp; English.
XX
CC The present sequence represents a SPLICE erbb-2 receptor protein. The
CC protein has an in-frame deletion of 16 amino acids, 2 of which are
CC conserved cysteine residues, compared to the unspliced protein. The
CC erbb-2 polynucleotide is used to construct probes for detecting
CC disorders of cell transformation such as cancer. Antibodies to the
CC protein may be used to detect SPLICE erbb-2 in a sample. Agents
CC (e.g. antisense oligonucleotides) which inhibit the expression of
CC SPLICE erbb-2 are useful for reducing tumor cell proliferation and
CC treating cancer. Substances which stimulate SPLICE erbb-2 are useful
CC for treating conditions involving damaged cells including conditions
CC in which degeneration of tissue occurs such as arthropathy, bone
CC resorption, inflammatory diseases, degenerative disorders of the
CC central nervous system and wound healing.
XX
SQ Sequence 1255 AA;
Query Match 98.0%; Score 6707; DB 21; Length 1255;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1240; Conservative 5; Mismatches 10; Indels 6; Gaps 2;
Qy 1 MELAALCRWGLLALLPPGAASQVCTGDMKRLRLPASPTHLDMLRHLYGCGVQVGNL 60
Db 1 MELAALCRWGLLALLPPGAASQVCTGDMKRLRLPASPTHLDMLRHLYGCGVQVGNL 60
Qy 61 ELYLPTNASLSFLQDIOEQVFNFTVSWFVRPKVSASHLEORLRIVRGTQLPEDNYAL 120
Db 61 ELYLPTNASLSFLQDIOEQVFNFTVSWFVRPKVSASHLEORLRIVRGTQLPEDNYAL 114
Qy 121 AVLNDGDPNNTPVTVGASPGGLREQLRLSLTELKGGVLIORNPOLCYQDTILWKDIFH 180
Db 115 AVLNDGDPNNTPVTVGASPGGLREQLRLSLTELKGGVLIORNPOLCYQDTILWKDIFH 174
Qy 181 KNNQALTLIDTNRSRACHPCSPMKGSRGSCWGSSEDCQSLTRTVTCAGGCARCKGPLPTD 240
Db 175 KNNQALTLIDTNRSRACHPCSPMKGSRGSCWGSSEDCQSLTRTVTCAGGCARCKGPLPTD 234
Qy 241 CCHQCAAGCTGPKGSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGAS 300
Db 235 CCHQCAAGCTGPKGSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGAS 294
Qy 301 CVTACPNYLSLTDVGSCTLVCPLNHOEVTAEDGTQRCCKSPCARVCYGLGMEHLREVR 360
Db 295 CVTACPNYLSLTDVGSCTLVCPLNHOEVTAEDGTQRCCKSPCARVCYGLGMEHLREVR 354
Qy 361 AVTSANIQEPAGCKKIFGSLAFPEPSDGPASNTAPLOPEQLQVFTLBEITGILYISA 420
Db 355 AVTSANIQEPAGCKKIFGSLAFPEPSDGPASNTAPLOPEQLQVFTLBEITGILYISA 414
Qy 421 WPDLSPLDSVFNQLQVIRGRILHNGAYSLSLQGLISWLGRLSLRELGSGLAIHHNTHL 480
Db 415 WPDLSPLDSVFNQLQVIRGRILHNGAYSLSLQGLISWLGRLSLRELGSGLAIHHNTHL 474
Qy 481 CFVHTVPWDQLFRNPHQALLHTANRDECEVCGGLACHQICARGHCWGPPTQCVNCSQF 540
Db 475 CFVHTVPWDQLFRNPHQALLHTANRDECEVCGGLACHQICARGHCWGPPTQCVNCSQF 534
Qy 541 LRQECVCECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFPGPADQCVACAHYKDDP 600
Db 535 LRQECVCECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFPGPADQCVACAHYKDDP 594
Qy 601 FCVARCPGSKPDLSPYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGPASPLTISI 660


```
Db 475 CFVHTVPMDQLFRNP HQALLHTANRDEDCVGEGLACHOLCARGHCWGPGPTQCVCNCSQF 534
Qy 541 LRQGEVVEECRVLOGLPREYNARHCLPCHPECPQNGSVTCFGEADQCACAHYKDDPP 600
Db 535 LRQGEVVEECRVLOGLPREYNARHCLPCHPECPQNGSVTCFGEADQCACAHYKDDPP 594
Qy 601 FCVARCPGKVPDLSYMPKWKFPDEEGACQPCPINCTHSCVDLDDKGCAPQASPLTSI 660
Db 595 FCVARCPGKVPDLSYMPKWKFPDEEGACQPCPINCTHSCVDLDDKGCAPQASPLTSI 654
Qy 661 VSAVVGILLVVLGVVFGILIKRQOKIRKYTMRLLOETELVPLTPSGAMPNQAQMRI 720
Db 655 ISAVVGILLVVLGVVFGILIKRQOKIRKYTMRLLOETELVPLTPSGAMPNQAQMRI 714
Qy 721 LKETELRKVKLGSGAFGVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVM 780
Db 715 LKETELRKVKLGSGAFGVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVM 774
Qy 781 AGVSPYVSRLLGICLTSTVOLVTOLMPYGCLLDHVRENRRGLSGQDLLNMCQIAKMS 840
Db 775 AGVSPYVSRLLGICLTSTVOLVTOLMPYGCLLDHVRENRRGLSGQDLLNMCQIAKMS 834
Qy 841 YLEDRVLRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALESI 900
Db 835 YLEDRVLRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALESI 894
Qy 901 LRRRFTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITDVYMI 960
Db 895 LRRRFTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITDVYMI 954
Qy 961 MVKCMIDSECRPRELVSFSEMRARDPQRFVVIQNEEDLGPASPLDSTFYRSLLEDDDM 1020
Db 955 MVKCMIDSECRPRELVSFSEMRARDPQRFVVIQNEEDLGPASPLDSTFYRSLLEDDDM 1014
Qy 1021 GDLVDAEYLVPOQGFCCPDPAAGAGGMVHRRSSSTRSGGDI.TLGLSEPEEAPRSP 1080
Db 1015 GDLVDAEYLVPOQGFCCPDPAAGAGGMVHRRSSSTRSGGDI.TLGLSEPEEAPRSP 1074
Qy 1081 LAPSEGAGSDVFDGLNGAAKGLQSLTHDPSPLOQRYSEDPVPLPSETGYVAPLTCS 1140
Db 1075 LAPSEGAGSDVFDGLNGAAKGLQSLTHDPSPLOQRYSEDPVPLPSETGYVAPLTCS 1134
Qy 1141 POPEVYNQDVRPQPPSPREGPLPAARPAAGATLERAKTLPKNGVWVDVAFGAVENP 1200
Db 1135 POPEVYNQDVRPQPPSPREGPLPAARPAAGATLERPKTLPKNGVWVDVAFGAVENP 1194
Qy 1201 EYLTPQGGAAAPQHPPPAFSPAFNLYYWDQDPPERGAAPPSTFKGTPTAENPEYGLDVP 1260
Db 1195 EYLTPQGGAAAPQHPPPAFSPAFNLYYWDQDPPERGAAPPSTFKGTPTAENPEYGLDVP 1254
Qy 1261 V 1261
Db 1255 V 1255

RESULT 11
AAE24067
ID AAE24067 standard; Protein; 1255 AA.
XX
AC AAE24067;
XX
DT 23-SEP-2002 (first entry)
XX
DE Human Her-2 protein.
XX
KW Human; Her-2; epidermal growth factor receptor 2; infection; cancer;
KW hyperproliferative disorder; prophylaxis; inflammation; antisense;
KW tumour; gene therapy; phosphorothioate backbone.
XX
OS Homo sapiens.
XX
PN WO200222636-A1.
XX
```

```
PD 21-MAR-2002.
XX
PF 12-SEP-2001; 2001WO-US28572.
XX
PR 15-SEP-2000; 2000US-0663834.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Bennett CF, Cowseart LM;
XX
DR WPI; 2002-471192/50.
DR N-PSDB; AAD38904.
XX
PT Novel antisense oligonucleotide which modulates the expression of Human
PT Epidermal Growth Factor receptor, Her2, is useful for treating tumors
PT Inflammation or to prevent infection in humans -
XX
PS Example 13; Page 95-107; 116pp; English.
XX
CC The invention relates to antisense compounds targetted to a nucleic
CC acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)
CC that specifically hybridises with and inhibits the expression of Her2.
CC Antisense compounds of the invention are used for treating diseases or
CC conditions associated with Her2 such as hyperproliferative disorders
CC e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,
CC neural or cardiac cancer. They are also useful prophylactically e.g.
CC to prevent or delay infection, inflammation and tumour formation. The
CC invention is also used in gene therapy. The present sequence is human
CC Her-2 protein.
XX
SQ Sequence 1255 AA;
Query Match 98.08; Score 6707; DB 23; Length 1255;
Best Local Similarity 98.34; Pred. No. 0;
Matches 1240; Conservative 5; Mismatches 10; Indels 6; Gaps 2;
Qy 1 MELAALCRWGLLALLPFGAASCTQCTGDMKRLPASPETHDMLRHLVGCQVQGNL 60
Db 1 MELAALCRWGLLALLPFGAASCTQCTGDMKRLPASPETHDMLRHLVGCQVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIOEVQFNFTVSFWLRVPKVSASHLEQRLRIVRGTLQFEDNYAL 120
Db 61 ELTYLPTNASLSFLQDIOEVQ--GYVLIAHQVRQVPL---QRLRIVRGTLQFEDNYAL 114
Qy 121 AVLDNGDPLNNTTPTVTCASPGGLRELOIRSLTEILKGGVLIQRPOLCYODTILWKDIFH 180
Db 115 AVLDNGDPLNNTTPTVTCASPGGLRELOIRSLTEILKGGVLIQRPOLCYODTILWKDIFH 174
Qy 181 KNNQALTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVCAAGCARKGPLEPTD 240
Db 175 KNNQALTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVCAAGCARKGPLEPTD 234
Qy 241 CCHQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRTYFGAS 300
Db 235 CCHQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRTYFGAS 294
Qy 301 CVTACPYNILSTDVGSCTLYCPLHNQVETAEDGTQCEKSKPCARVCYGLGMEHLREVR 360
Db 295 CVTACPYNILSTDVGSCTLYCPLHNQVETAEDGTQCEKSKPCARVCYGLGMEHLREVR 354
Qy 361 AVTSANTQEFAGCKKIFGSLAFIPESFDGDPASNTAPLQEQVLFTEITGYLYISA 420
Db 355 AVTSANTQEFAGCKKIFGSLAFIPESFDGDPASNTAPLQEQVLFTEITGYLYISA 414
Qy 421 WPDSLPLDSVFNQLQVIRGRILHNGAYSLLTQGLGISWGLRSLRSLRELSGLALIHNNTHL 480
Db 415 WPDSLPLDSVFNQLQVIRGRILHNGAYSLLTQGLGISWGLRSLRSLRELSGLALIHNNTHL 474
Qy 481 CFVHTVPMDQLFRNP HQALLHTANRDEDCVGEGLACHOLCARGHCWGPGPTQCVCNCSQF 540
Db 475 CFVHTVPMDQLFRNP HQALLHTANRDEDCVGEGLACHOLCARGHCWGPGPTQCVCNCSQF 534
Qy 541 LRQGEVVEECRVLOGLPREYNARHCLPCHPECPQNGSVTCFGEADQCACAHYKDDPP 600
```

Db	535	LRGQCEVBEQVQLGLPREYVNAHCLPCHPEQCPQNGSVTCFGEADQCACAHYKDDP	594
Qy	601	FCVARCPSGVKPDLSYMPIWKFPPDEGACQPCPINCTHSCVDDLDKQCPAQRASPLTSI	660
Db	595	FCVARCPSGVKPDLSYMPIWKFPPDEGACQPCPINCTHSCVDDLDKQCPAQRASPLTSI	654
Qy	661	VSADVGLLVVVLGVVFGILIKRQOKIRKYTMRLLOTELVEPLTPSGAMPNQAMRI	720
Db	655	ISAVVGILLVVVLGVVFGILIKRQOKIRKYTMRLLOTELVEPLTPSGAMPNQAMRI	714
Qy	721	LKETELRKVKVLGSGAFGVYKGIWIPGENVKIPVAIKVIRENTSPKANKEIIDEAYVM	780
Db	715	LKETELRKVKVLGSGAFGVYKGIWIPGENVKIPVAIKVIRENTSPKANKEIIDEAYVM	774
Qy	781	AGVSPYVSRLLIGLCTSTVOLVTPQMPYGCILLDHVRENRRGLSODLLNMCQIAKMS	840
Db	775	AGVSPYVSRLLIGLCTSTVOLVTPQMPYGCILLDHVRENRRGLSODLLNMCQIAKMS	834
Qy	841	YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKWMALESI	900
Db	835	YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKWMALESI	894
Qy	901	LRRRPTHOSDVMSYGVTVWELMTGAKYDGIIPAREIPDLLEKGERLPQPPICITDVYMI	960
Db	895	LRRRPTHOSDVMSYGVTVWELMTGAKYDGIIPAREIPDLLEKGERLPQPPICITDVYMI	954
Qy	961	MVKCWMIDSECRPRELVSEFSSRMARDPQRFVVIQNEEDLGPASPLDSTFYRSILLEDDDM	1020
Db	955	MVKCWMIDSECRPRELVSEFSSRMARDPQRFVVIQNEEDLGPASPLDSTFYRSILLEDDDM	1014
Qy	1021	GDLVDAEYLVPOQGFPCDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEBEAPRSP	1080
Db	1015	GDLVDAEYLVPOQGFPCDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEBEAPRSP	1074
Qy	1081	LAPSEGAGSDVFDGLGMAKGLQSLTHDPSPLORYSEPTVPLPSETDGYVAPLTCS	1140
Db	1075	LAPSEGAGSDVFDGLGMAKGLQSLTHDPSPLORYSEPTVPLPSETDGYVAPLTCS	1134
Qy	1141	POPEYVNOQVVRPOPSPREGPLPAARPAAGATLERAKTSLSPKNGVVKDVFAPGAVENP	1200
Db	1135	POPEYVNOQVVRPOPSPREGPLPAARPAAGATLERAKTSLSPKNGVVKDVFAPGAVENP	1194
Qy	1201	EYLTPOGGAAPQHPHPPAFSPAFNLVYWDQDPPERGAPPSTFKGTPPTAENPEYLGLDVP	1260
Db	1195	EYLTPOGGAAPQHPHPPAFSPAFNLVYWDQDPPERGAPPSTFKGTPPTAENPEYLGLDVP	1254
Qy	1261	V	1261
Db	1255	V	1255
RESULT 12			
AAE20479			
ID	AAE20479 standard; Protein; 1255 AA.		
XX			
AC	AAE20479;		
XX			
DT	01-JUL-2002 (first entry)		
XX			
DE	Human Her-2/neu protein.		
XX			
KW	Human; Her-2/Neu protein; immune response; gene therapy; breast cancer;		
XX			
KW	human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.		
XX			
OS	Homo sapiens.		
XX			
FH	Key		
FT	Location/Qualifiers		
FT	1021..1030		
FT	/note= "Naturally processed HLA-B44-restricted epitope"		
XX			
PN	WO200214503-A2.		
XX			

PD	21-FEB-2002.	
XX		
PF	14-AUG-2001; 2001WO-US41733.	
XX		
PR	14-AUG-2000; 2000US-225152P.	
PR	28-SEP-2000; 2000US-236428P.	
PR	21-FEB-2001; 2001US-270520P.	
XX		
PA	(CORI-) CORIXA CORP.	
XX		
PI	Hand-zimmermann S, Cheever MA, Foy TM, Lodes MJ, Kalos MD;	
PI	Mcneill PD, Vedvick TS;	
XX		
DR	WPI; 2002-280758/32.	
DR	N-PSDB; AAD32743.	
XX		
PT	Novel isolated Her-2/Neu polypeptide composition useful for therapy,	
PT	prevention and diagnosis of cancer, preferably breast cancer -	
XX		
PS	Disclosure; Page 114-117; 129pp; English.	
XX		
CC	The invention relates to an isolated Her-2/Neu polypeptide composition effective for eliciting an immune response. The invention is useful for eliciting an immune response in a patient, where the patient is human leukocyte antigen (HLA)-B44 positive or is affected with breast cancer. The composition is useful for the therapy and diagnosis of cancer, preferably breast cancer, in pharmaceutical compositions, e.g., vaccine and other compositions for the diagnosis, prevention and treatment of human malignancies for stimulating and/or expanding T cells specific for Her-2/Neu polypeptide and for inhibiting the development of cancer in a human patient, as probe or primer for nucleic acid hybridisation, to selectively form duplex molecules with complementary stretches of the entire Her-2/Neu gene or gene fragments of interest to isolate a full length gene from a suitable library, and to direct expression of a polypeptide in appropriate host cells. The composition is useful in prophylactic or therapeutic applications and for the treatment of cancer, preferably for the immunotherapy of breast cancer and other Her-2/Neu-associated malignancies. The invention is useful in gene therapy. The present sequence is human Her-2/neu protein.	
XX		
SQ	Sequence	1255 AA;
Query Match		
Best Local Similarity 98.0%; Score 6707; DB 23; Length 1255;		
Matches 1240; Conservative 5; Mismatches 10; Indels 6; Gaps 2;		
Qy	1	MELAAALCRWGLLLALLPPGAASTQVCTGDMKRLPASPETHLDMLRHLYGCGVQGNL 60
Db	1	MELAAALCRWGLLLALLPPGAASTQVCTGDMKRLPASPETHLDMLRHLYGCGVQGNL 60
Qy	61	ELTYLPTNASLSFLQDIOEVQFNFTVSFWLRVPKVSASHLEORLVRGTOLPEDNYAL 120
Db	61	ELTYLPTNASLSFLQDIOEVQ--GYVLIANHQRQVPL-----QRLRVRGTQFEDNYAL 114
Qy	121	AVLDNGDPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFH 180
Db	115	AVLDNGDPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFH 174
Qy	181	KNNQALATLIDTNRSRACHPCSPMKGSRCWGSSSDCQSLTRTVACGACRCKGPLPTD 240
Db	175	KNNQALATLIDTNRSRACHPCSPMKGSRCWGSSSDCQSLTRTVACGACRCKGPLPTD 234
Qy	241	CCHEOCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGAS 300
Db	235	CCHEOCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGAS 294
Qy	301	CVTACPNYLSLTDVGSCTLLVCPLNHNOEVTADGTQRCCKSKPCARVCYGLGHEHLREVR 360
Db	295	CVTACPNYLSLTDVGSCTLLVCPLNHNOEVTADGTQRCCKSKPCARVCYGLGHEHLREVR 354
Qy	361	AVTSANIQEFAGCKKIFGSLAFLPESFGDGPASNTAPLOPEQLQVFETLEETGYLYISA 420

Db 355 AVTSANIQFAGCKKIFGSLAFAPESFDGDPASNTAPLOPQLOVFETLEBITGYLYISA 414
Qy 421 WPSLPLDSVFQNLQVIRGRILHNGAYSLTIQGLGISWLGRLSRELGSGLALHNNTHL 480
Db 415 WPSLPLDSVFQNLQVIRGRILHNGAYSLTIQGLGISWLGRLSRELGSGLALHNNTHL 474
Qy 481 CFVHTVPMQDLFRPNPHQALLHTANRPEDECVEGEGACHOLCARGHCWGPGTQCVCNSQF 540
Db 475 CFVHTVPMQDLFRPNPHQALLHTANRPEDECVEGEGACHOLCARGHCWGPGTQCVCNSQF 534
Qy 541 LRGOECVEECRVLOGLPREYVNAHCLPCHPECOQNGSVTCFGEADQCACAHYKDP 600
Db 535 LRGOECVEECRVLOGLPREYVNAHCLPCHPECOQNGSVTCFGEADQCACAHYKDP 594
Qy 601 FCVARCPSGVKPDLSYMPIWKFPPDEGACQPCPNCTHSCVDLDDKGCAPAEORASPLTSI 660
Db 595 FCVARCPSGVKPDLSYMPIWKFPPDEGACQPCPNCTHSCVDLDDKGCAPAEORASPLTSI 654
Qy 661 VSAVVGLLVVVGWVFGILIKRQOKIRKTKYMRLLQETELVEPLTPSGAMPNQAOWRI 720
Db 655 ISAVVGLLVVVGWVFGILIKRQOKIRKTKYMRLLQETELVEPLTPSGAMPNQAOWRI 714
Qy 721 LKETELRKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEIILDEAYVM 780
Db 715 LKETELRKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEIILDEAYVM 774
Qy 781 AGVSPYVSRLLGLTSTVQLMYPYGLLDHVRNRRGLSGQDLNNCMQIAKMS 840
Db 775 AGVSPYVSRLLGLTSTVQLMYPYGLLDHVRNRRGLSGQDLNNCMQIAKMS 834
Qy 841 YLEDVRLVHRDLAARNVLKSPNVHKITDFGLARLLDIDETEHADGKVPKKNWALESI 900
Db 835 YLEDVRLVHRDLAARNVLKSPNVHKITDFGLARLLDIDETEHADGKVPKKNWALESI 894
Qy 901 LRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPPQPICTIDVYMI 960
Db 895 LRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPPQPICTIDVYMI 954
Qy 961 MVKWMIDSECRPRELVSFESRMARDPQRFVVIQNEGLGASPLDSTFYRSLLEDMD 1020
Db 955 MVKWMIDSECRPRELVSFESRMARDPQRFVVIQNEGLGASPLDSTFYRSLLEDMD 1014
Qy 1021 GDLVDAEYLVPOQGFCCPDAPGAGGMVHRHRSSTRSGGDLTLGLEPSEEAAPRSP 1080
Db 1015 GDLVDAEYLVPOQGFCCPDAPGAGGMVHRHRSSTRSGGDLTLGLEPSEEAAPRSP 1074
Qy 1081 LAPSEGAGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPVPLPSETDGYVAPLTC 1140
Db 1075 LAPSEGAGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPVPLPSETDGYVAPLTC 1134
Qy 1141 POPEYVNPQDVRPQPPSPREGPLPAARAGATLERAKTSLSPGKNGVVKDVFAGGAVENP 1200
Db 1135 POPEYVNPQDVRPQPPSPREGPLPAARAGATLERAKTSLSPGKNGVVKDVFAGGAVENP 1194
Qy 1201 EYLTPOGGAAPQHPHPPAFSADFNLXYWDDPPERGAPPSTFKTPTAENPEYGLDVP 1260
Db 1195 EYLTPOGGAAPQHPHPPAFSADFNLXYWDDPPERGAPPSTFKTPTAENPEYGLDVP 1254
Qy 1261 V 1261
Db 1255 V 1255

RESULT 13

AA51143
ID AA51143 standard; Protein; 1255 AA.XX
AC AA51143;

XX 17-JUN-2002 (first entry)

XX Human Her-2/neu oncogene-encoded p185 glycoprotein.

KW Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;
tyrosine kinase; receptor; c-erbB2; gene therapy.

XX Homo sapiens.

Key Location/Qualifiers
Domain 1..653
/note= "extracellular domain"
Domain 676..1255
/note= "intracellular domain"
Domain 990..1255
/note= "phosphorylation domain"

W0200212341-A2.

14-FEB-2002.

03-AUG-2001; 2001WO-US24283.

03-AUG-2000; 2000US-0632507.

(CORI-) CORIXA CORP.
(SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

Cheever MA, Gheysen D;

WPI; 2002-241743/29.

N-PSDB; ABA92250.

Her-2/neu fusion protein for treating or preventing cancer by eliciting
or enhancing an immune response to the protein, has Her-2/neu
extracellular domain fused to Her-2/neu intracellular or
phosphorylation domain

Claim 68; Fig 7; 141pp; English.

The present sequence is that of human Her-2/neu (p185 glycoprotein or c-erbB2). An oncogenic self-protein and target for anti-cancer vaccines. The Her-2/neu gene is amplified and p185 is overexpressed in a variety of cancers, including breast, ovarian, colon, lung and prostate cancer. Her-2/neu is a member of the tyrosine kinase family of receptor-like glycoproteins. It comprises an extracellular domain with homology to the epidermal growth factor receptor (EGFR), a highly hydrophobic transmembrane domain and a C-terminal intracellular domain that also shows homology to EGFR. Its overexpression correlates with a poor prognosis in breast and ovarian cancers. The invention provides Her-2/neu fusion proteins, nucleic acids encoding them, viral vectors, and vaccines comprising the fusion proteins or nucleic acid molecules. In preferred fusion proteins, the extracellular domain of a Her-2/neu protein is fused to a Her-2/neu intracellular domain or phosphorylation domain (or its beta-fragment). An immune response to Her-2/neu protein is elicited or enhanced by administering the fusion protein in the form of a vaccine, or by transfecting cells of an animal *ex vivo* with a nucleic acid encoding the fusion protein, and delivering the transfected cells to the animal. The fusion proteins, nucleic acids, and isolated specific T-cells are useful for inhibiting the development of a cancer, especially breast, ovarian, colon, lung or prostate cancer in a patient. T cells that specifically react with a Her-2/neu fusion protein can be used to remove tumour cells from a sample in order to inhibit the development of cancer in a patient.

XX Sequence 1255 AA;

Query Match 98.0%; Score 6707; DB 23; Length 1255;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1240; Conservative 5; Mismatches 10; Indels 6; Gaps 2;

Qy 1 MELAALCRWGLLLALLPFGAASCTGCTDKMLRPLASPETHLDMRLHYGCGVQGNL 60

Db 1 MELAALCRWGLLLALLPFGAASCTGCTDKMLRPLASPETHLDMRLHYGCGVQGNL 60

Db 1135 POEYVNPQDVRPOPPSPREGPLPAARAGATLERPKTSLSPGKNGVVKDVFAFGGAVENP 1194
Qy 1201 EYLTPOGGAAPQHPHPPAFSPADNLYWDODPERGAPBSTFKGTPTAENPEYLGLDVP 1260
Db 1195 EYLTPOGGAAPQHPHPPAFSPADNLYWDODPERGAPBSTFKGTPTAENPEYLGLDVP 1254
Qy 1261 V 1261
Db 1255 V 1255
RESULT 15
AAR39568
ID AAR39568 standard; Protein; 1433 AA.
XX AAR39568;
DT 07-FEB-1994 (first entry)
XX Sequence of c-erbB-2 tumour antigen.
XX Tumour antigen; c-erbB-2; glycoprotein.
XX Homo sapiens.
XX WO9316185-A.
XX 19-AUG-1993.
XX 05-FEB-1993; 93WO-US01055.
XX 06-FEB-1992; 92US-0831967.
XX (CETU) CETUS ONCOLOGY CORP.
XX (CREA-) CREATIVE BIOMOLECULES INC.
XX Houston LL, Huston JS, Oppermann H, Ring DB;
XX WPI; 1993-272889/34.
XX N-PSDB; AAQ46083.
XX New single chain Fv polypeptide binding to C-erbB-2 tumour
XX antigen - for imaging or treating breast or ovarian cancer etc.
XX Disclosure; pages 48-54; 87pp; English.
XX c-erbB-2 refers to a protein antigen expressed on the surface of
XX tumour cells, such as breast and ovarian tumour cells, which is an
XX approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric
XX pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39583 represents
XX the location of a stop codon in AAQ46083.
SQ Sequence 1433 AA;
Query Match 97.3%; Score 6664; DB 14; Length 1433;
Best Local Similarity 97.8%; Pred. No. 0;
Matches 1233; Conservative 7; Mismatches 15; Indels 6; Gaps 2;
Qy 1 MELAALCRWGLLALLPPGAAGTQCTGDMKRLPASPETHLDMRLHLYQCVQVQGNL 60
Db 1 MELAALCRWGLLALLPPGAAGTQCTGDMKRLPASPETHLDMRLHLYQCVQVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIQEVO---GYVLIANQVRQVPL---QRLRIVRGQLFEDNYAL 120
Db 61 ELTYLPTNASLSFLQDIQEVO---GYVLIANQVRQVPL---QRLRIVRGQLFEDNYAL 114
Qy 121 AVLNDGPDPLNNTTPTVTCASPGGLRELQSLTEILKGGVLIQNPOLCYQDTILWKDIFH 180
Db 115 AVLNDGPDPLNNTTPTVTCASPGGLRELQSLTEILKGGVLIQNPOLCYQDTILWKDIFH 174
Qy 181 KNNQALTLIDTNRSRACHPCSPMKGSRGWESSEDCQSLTRTVACAGGCARCKGPLPTD 240
Db 175 KNNQALTLIDTNRSRACHPCSPMKGSRGWESSEDCQSLTRTVACAGGCARCKGPLPTD 234

Db 61 ELTYLPTNASLSFLQDIQEVO---GYVLIANQVRQVPL---QRLRIVRGQLFEDNYAL 114
Qy 121 AVLNDGPDPLNNTTPTVTCASPGGLRELQSLTEILKGGVLIQNPOLCYQDTILWKDIFH 180
Db 115 AVLNDGPDPLNNTTPTVTCASPGGLRELQSLTEILKGGVLIQNPOLCYQDTILWKDIFH 174
Qy 181 KNNQALTLIDTNRSRACHPCSPMKGSRGWESSEDCQSLTRTVACAGGCARCKGPLPTD 240
Db 175 KNNQALTLIDTNRSRACHPCSPMKGSRGWESSEDCQSLTRTVACAGGCARCKGPLPTD 234
Qy 241 CCHQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDIFESMPNPEGRYTFGAS 300
Db 235 CCHQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDIFESMPNPEGRYTFGAS 294
Qy 301 CVTACPNYLTSTVGSCITLVCPLNHNEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVR 360
Db 295 CVTACPNYLTSTVGSCITLVCPLNHNEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVR 354
Qy 361 AVTSANTOEFAGCKKIFGSLAFLESFDGDPASNTAPLOPELOVFEITGLYLYISA 420
Db 355 AVTSANTOEFAGCKKIFGSLAFLESFDGDPASNTAPLOPELOVFEITGLYLYISA 414
Qy 421 WPDLSLPDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRELGSGLAIHNNTHL 480
Db 415 WPDLSLPDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRELGSGLAIHNNTHL 474
Qy 481 CFVHTVPMQDLFRNPHQALLTANRPEDECVGEGGLACHOLCARGHCWGPGTQCVNCSQF 540
Db 475 CFVHTVPMQDLFRNPHQALLTANRPEDECVGEGGLACHOLCARGHCWGPGTQCVNCSQF 534
Qy 541 LRQGECEVCEVLQGLPREYNARHCLPCHPECOPQNGSVTCFGEADQCVACAHYKDDP 600
Db 535 LRQGECEVCEVLQGLPREYNARHCLPCHPECOPQNGSVTCFGEADQCVACAHYKDDP 594
Qy 601 FCVARCSPGVKPDLSYMPIMKFPDEGACQPCPINCTHSCVDLDDKGPAPQASPLTSI 660
Db 595 FCVARCSPGVKPDLSYMPIMKFPDEGACQPCPINCTHSCVDLDDKGPAPQASPLTSI 654
Qy 661 VSAVVGILLVVLGVVFGILIKRQOKIRKYMRLLOETELVPLPLTPSGAMPNQAOMRI 720
Db 655 ISAVVGILLVVLGVVFGILIKRQOKIRKYMRLLOETELVPLPLTPSGAMPNQAOMRI 714
Qy 721 LKETELRKVKVLSGAFCTVVGKIWDGPNVKIPVAIKVRENTSPKANKEILDEAVYM 780
Db 715 LKETELRKVKVLSGAFCTVVGKIWDGPNVKIPVAIKVRENTSPKANKEILDEAVYM 774
Qy 781 AGVSPYVSRLIGLICLTSTVQLVTLQMPYGCILLDHVRENRLGSGQDILLNMCQIAKWS 840
Db 775 AGVSPYVSRLIGLICLTSTVQLVTLQMPYGCILLDHVRENRLGSGQDILLNMCQIAKWS 834
Qy 841 YLEDRVLRDLAARNVLKSPNHVKITDPLGLARLLDIDETEHADGKVPKIMWALESI 900
Db 835 YLEDRVLRDLAARNVLKSPNHVKITDPLGLARLLDIDETEHADGKVPKIMWALESI 894
Qy 901 LRRRPTHOSDVSXGVTVWELMTCAKPYDGIIPAREIPDLLEKGERLPQPPICITDVMYI 960
Db 895 LRRRPTHOSDVSXGVTVWELMTCAKPYDGIIPAREIPDLLEKGERLPQPPICITDVMYI 954
Qy 961 MVKCMWIDSECRPRELVSFBSRMARDPQRFVITQNEEDLGPASPLDSTFYRSLLEDDDM 1020
Db 955 MVKCMWIDSECRPRELVSFBSRMARDPQRFVITQNEEDLGPASPLDSTFYRSLLEDDDM 1014
Qy 1021 GDLVDAEYLVPOQGFCCPDAPAGAGMVHHRHSSSTRSGGDLITLGLPSESEAPRSP 1080
Db 1015 GDLVDAEYLVPOQGFCCPDAPAGAGMVHHRHSSSTRSGGDLITLGLPSESEAPRSP 1074
Qy 1081 LAPSEGAGSDVFDGLGMAKGLQSLPDPSPLOQRYSEDPVPLPSETDGYVAPLTC 1140
Db 1075 LAPSEGAGSDVFDGLGMAKGLQSLPDPSPLOQRYSEDPVPLPSETDGYVAPLTC 1134
Qy 1141 POEYVNPQDVRPOPPSPREGPLPAARAGATLERAKTSLSPGKNGVVKDVFAFGGAVENP 1200

Search completed: July 22, 2003, 08:40:22
Job time : 43.1589 secs

QY	241	CCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTPESMNPDEGRYTFGAS	300
Db	235		
QY	301	CVTACPNYVLTVDGSCCTLVCPHNOEVTAEQTCCKKPCARVCYGLGMEHLREVR	360
Db	295		
QY	361	AVTSANIQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEBITGYLYISA	420
Db	355		
QY	421	WPDLSPLDSVFQNLQVIRGRILHNGAYSILTLQGLGISWGLRSRELGSGLALIHNTHL	480
Db	415		
QY	481	CFVHTVPWDQLFRNPQHALLHTANRDECECVGEGACHQLCARGHCWGPGTQCVCNCSQF	540
Db	475		
QY	541	LRGQECVEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDDP	600
Db	535		
QY	601	FCVARCPGVPDLSYMPIWKPDEEGACQPCPCINCTHSCVDLDDKGCAPQORASPLTSI	660
Db	595		
QY	661	VSAVVGILLVVVLGVVFGILIKRQOKIRKYMRLLOETELVEPLTPSGAMPNQAQMRI	720
Db	655		
QY	721	LKETELRKVKVLGSGAFVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVM	780
Db	715		
QY	781	AGVSPYVSRLLGICLTSTVLTQMPYGCCLDHVRENRLGSDQLLNWCMIAGMS	840
Db	775		
QY	841	YLEDVRLVHRDLAARNLVKSPNHVKITDFGLARLLIDETEHADGKVPKWMALESI	900
Db	835		
QY	901	LRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYMI	960
Db	895		
QY	961	MVKCWMIDSECRPRFRELVSFERSMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDM	1020
Db	955		
QY	1021	GDLDVAEEYLPQOGFCFDPAPGAGGVHHRHSSSTRSCGGDLTLGLEPSEEAAPRSP	1080
Db	1015		
QY	1081	LAPSEGAGSDVFDGDLGNAKGLQSLPTHDPSPLOQRYSEDPVPLSETDGYVAPLTC	1140
Db	1075		
QY	1141	POPEYVNOPDVRPOPSPREGPLPAARPAAGATLERAKTLPSPKNGVVKDVFAGGAVENP	1200
Db	1135		
QY	1201	EYLTPOGGAAQPHPPPAFPAFNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVP	1260
Db	1195		
QY	1261	V 1261	
Db	1255	V 1255	

THIS PAGE BLANK (USPTO)

Nature 309, 270-273, 1984
A;Title: ATP-stimulated interaction between epidermal growth factor receptor and superoxide dismutase
A;Reference number: A38023; MUID:84191554; PMID:6325948
A;Contents: annotation; receptor activity
A;Note: the EGF receptor (and other tyrosine kinases) can nick double-stranded DNA
R;Chen, W.S.; Lazar, C.S.; Lund, K.A.; Welsh, J.B.; Chang, C.P.; Walton, G.M.; Der, C.J.; Cell 59, 33-43, 1989
A;Title: Functional independence of the epidermal growth factor receptor from a domain x
A;Reference number: A3331; MUID:90003233; PMID:2790960
A;Contents: annotation; internalization signal
C;Comment: Binding of EGF to the receptor leads to internalization of the EGF-receptor
C;Genetics:
A;Gene: GDB:EGFR
A;Cross-references: GDB:120610; OMIM:131550
A;Map position: 7p12.3-7p12.1
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosphatase
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-1210/Product: EGF receptor #status predicted <MAT>
F;25-645/Domain: extracellular #status predicted <EXT>
F;75-300/Domain: EGF receptor extracellular domain repeat <BE1>
F;390-600/Domain: EGF receptor extracellular domain repeat <BE2>
F;646-668/Domain: transmembrane #status predicted <TM>
F;669-1210/Domain: intracellular #status predicted <INT>
F;710-975/Domain: protein kinase homology <KIN>
F;718-726/Region: protein kinase ATP-binding motif
F;999-1046/Region: coated-pit mediated internalization signal
F;1047-1210/Region: inhibitory
F;128,175,352,413,444,528,603/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;745/Active site: Lys #status experimental

Query Match 46.0%; Score 3132; DB 1; Length 1210;
Best Local Similarity 49.1%; Pred. No. 9.6e-121;
Matches 626; Conservative 178; Mismatches 344; Indels 128; Gaps 22;

QY 11 LLAALLPPGAA-----STQVCTFNFTVFWLRVVKVSHLELYOGCOVQ 57
DB 14 LLAALCPASALBEKKVCOGTSNKLTLQTLGFEDHFLSL-----QRMFNCEVL 62
QY 58 GNLIELYLPNTNALSFLQDQVGVYVLIHNOVROVPLORLIRIVRGTQLFEDNYALVL 117
DB 63 GNLIELTVQRYNDLSFLKTLQEVAGYVLIANTVERIPLENQLIIRGNMYENSALVL 122
QY 118 DNGDPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILKWDIFKNN 177
DB 123 SNYD-----ANKTGLKELPMRNQLBELHGVAFSNPALCNVESIQWRDIVSSDF 172
QY 178 QLALTLIDTRSRACHPCSPMKGSRGWGESSEDCQSLTRTVACGGCA-RCKGPLPTDCC 236
DB 173 LSNMSMDFOHNLGSCQKDCSPNGSCWAGAEENCQKLTIKIICAAQCCSGRCRGSPPDCC 232
QY 237 HEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVYNTDTFESMPNPEGRYTFEGASCV 296
DB 233 HNQCAAGCTGPRSDCLVCKFRDEATCKTCTPPMLYNTPTTYQMDVNPBGKYSFGATCV 292
QY 297 TACPYNYLSTDVSGCTLVCLPHNQEVTAEDGTQRCCKSPCARVCYGLGMEHLREVRV 356
DB 293 KKCPRNVVTDHGSVCVRACGADSYEM-EEDGVKCKKCEGPKRCVKVNGIGIGEFKDSLSI 351
QY 357 TSANIQEFACGKIFGSLAFPLPSFDGDPASNTAPLOEQLOVPETLEETGYLYISAMP 416
DB 352 NATNIKHFKNCTSGDLHLPLVAFRGDSFTHPTPLDPQLBELDLIKTVKEITGFLLIQAWP 411
QY 417 DSLPLDSVFONLQVIRGILRHNGAYSILTLQGLGISWGLRSLRGLSGSLALIHNTLCLF 476
DB 412 ENRTDLHAFENLEIRGTQKHQGFSLAVSLNLSITLSGLRSLKSEISDGVIIISGNKLVCY 471
QY 477 VHTVPVQDLFRNPQALLHTANRDEDCVGEGLACHOLCARHGWCGPGPTQCVCNCSQFLR 536
DB 472 ANTINMKKLGTSQKTKIISNRGENSKATGQVCHALCSPEGCWGPEDPCVSCRNVSR 531
QY 537 GQECVBEICRVQLPRYVYVNRHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPC 596
DB 537 GQECVBEICRVQLPRYVYVNRHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPC 596

DB 532 GRSCVCKKLLGEPRFVENSICIOHPPELCPQAMNITCTGRGPNCIOCAHYIDGPHC 591
QY 597 VARCPGSKPDLGYMPTWKPDPDEGACQPCIPNCTHSCVDLDDKGCPAEQRASPLTSIVS 656
DB 592 VKTCPAGVGENNTL-VWKYADAGHVCHLCHPNTCYGCTPGLEGCTNGPKIP-SIAT 648
QY 657 AVVG-----ILLVVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNOAMR 713
DB 649 GMVGAALLLVVALGIG--LFMRHRRHVRKTLRLRLQLRELVEPLTPSGEAPNQALLR 705
QY 714 ILKETELRVKVLGSGAGFVYKGIWIPDGENVKIPVAIKVLRNTSPKANKELDEAYV 773
DB 706 ILKETEFKIKVLGSGAGFVYKGIWIPDGENVKIPVAIKELREATSPKANKELDEAYV 765
QY 774 MAGVGSPYVSRLLGICITSTVQLVTLQMPYGCCLLDHVRNRLGSGDGLNWCQIAKGM 833
DB 766 MASVDNPHVCRLLGICITSTVQLITQMPFGCLLDYVREHKDNIGSQYLLNWCQIAKGM 825
QY 834 SYLEDVRLVHRDLAARNVLKSNHVKITDFGLARLLDIDETEHADGGKVPKIMMALES 893
DB 826 NYLEDRLVHRDLAARNVLKTPQHVKITDFGLAKLIGABEKEYHAEKGGKVPKIMMALES 885
QY 894 ILRRRTHGSDVMSYGVYVWELMTFGAKPYDGIPIABEIPDLLEKGERLPQPPCTIDVYM 953
DB 886 ILHRIYTHGSDVMSYGVYVWELMTFGSKPYDGIPIABEIPDLLEKGERLPQPPCTIDVYM 945
QY 954 IMVKCMIDSECRPRELSEFSEFARMARDPQRFWVIQ-NEDLGPASPLDSTFYRSLLEDD 1012
DB 946 IMVKCMIDADSPKRELIIESEKWARDPQRLVYVQGDHMLPSTDFSNFYALMDEE 1005
QY 1013 DMGDLVDAEYLVPOQGFPCDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEERAPR 1072
DB 1006 DMDDVDVDAEYLVPOQGF-----SSPSTSR 1031
QY 1073 SPLAPSEAGSDVFDGDLGMAAKGLQSLTHDPSPLORYSEDPVPLPSET-DGVVAP 1130
DB 1032 TPLLSLSLSTN--NSTVACIDRNLQSCIKEDSFLORYSSDPTGALTEDSIDDTFL-- 1087
QY 1131 LTCSPQPEYVNOPDVRPQPPSPREGPLPAARPAATLERAKTLSPGKNGVVKVDFAFGGA 1190
DB 1088 -----PVPEYINQ-SVPRPAGSVONPVYHNOPLNP-----APSRDPHYQD--PHSTA 1132
QY 1191 VENPEYL-TPQGAAPQPHPPPAFPADNLYYWDQ-----DP-----PERGAP 1233
DB 1133 VGNPEYLVNTVQ-----PTCVNSTFDSPAHWAQKSGHQSILDNPDYQDFFPKAKP 1183
QY 1234 PSTFKGTPIHNPYEL 1249
DB 1184 NGIFKGS-TAENAYL 1198

RESULT 5
A53183
epidermal growth factor receptor precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 06-Jan-1995 #sequence revision 06-Jan-1995 #text change 18-Jun-1999
C;Accession: A53183; A43818; S24942; A28941; S45325; I49543
R;Luetteke, N.C.; Phillips, H.K.; Qiu, T.H.; Copeland, N.G.; Earp, H.S.; Jenkins, N.A.;
Genes Dev. 8, 399-413, 1994
A;Title: The mouse waved-2 phenotype results from a point mutation in the EGF receptor
A;Reference number: A53183; MUID:94170986; PMID:8125255
A;Accession: A53183
A;Molecule type: mRNA
A;Residues: 1-1210 <LUE>
A;Cross-references: GB:003425
R;Avivi, A.; Lax, I.; Ullrich, A.; Schlessinger, J.; Givol, D.; Morse, B.
Oncogene 6, 673-676, 1991
A;Title: Comparison of EGF receptor sequences as a guide to study the ligand binding site
A;Reference number: A43818; MUID:91232866; PMID:2030916
A;Accession: A43818
A;Molecule type: mRNA
A;Residues: 1-714 <AVI>
A;Cross-references: GB:X59698

R:Eisinger, D.P.; Serrero, G.
 submitted to the EMBL Data Library, June 1992
 A:Reference number: S24942
 A:Accession: S24942
 A:Molecule type: mRNA
 A:Residues: 969-971, 'K', 973-1115, 'D' <EIS>
 A:Cross-references: EMBL:12608
 R:Heisermann, G.J.; Gill, G.N.
 J. Biol. Chem. 263, 13152-13158, 1988
 A:Title: Epidermal growth factor receptor threonine and serine residues phosphorylated
 A:Reference number: A28941; MUID:88330814; PMID:3138223
 A:Accession: A28941
 A:Molecule type: protein
 A:Residues: 689-694, 'X', 696-704, 'L', 706-707, 989-992, 'XX', 995-996, 'X', 998-1000, 1002-1009,
 R:Hibbs, M.L.; Dunn, A.R.; Alexander, W.S.
 submitted to the EMBL Data Library, April 1994
 A:Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor
 A:Reference number: S45325
 A:Accession: S45325
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-971, 'K', 973-1210 <RES>
 A:Cross-references: EMBL:X78987; NID:9488830; PIDN:CAA55587.1; PID:9488831
 R:Paria, B.C.; Das, S.K.; Andrews, G.K.; Dey, S.K.
 Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993
 A:Title: Expression of the epidermal growth factor receptor gene is regulated in mouse h
 A:Reference number: 149643; MUID:93128380; PMID:7678348
 A:Accession: 149643
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 12-20, 22-132 <RES>
 A:Cross-references: GB:L06864; NID:g193001; PIDN:AAA53029.1; PID:g567201
 C:Genetics:
 A:Gene: EGFR
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphop
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:648-670/Domain: transmembrane #status predicted <TM>
 F:712-977/Domain: protein kinase homology <KIN>
 F:720-728/Region: protein kinase ATP-binding motif
 F:680,695/Binding site: phosphate (Thr) (covalent) #status experimental
 F:697,1070,1071/Binding site: phosphate (Ser) (covalent) #status experimental
 F:993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental
 F:1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental
 F:1197/Binding site: phosphate (Tyr) (covalent) #status experimental

Query Match 45.7%; Score 3111; DB 2; Length 1210;
 Best Local Similarity 49.1%; Pred. No. 6.9e-120;
 Matches 630; Conservative 169; Mismatches 352; Indels 132; Gaps 24;

Qy 11 LLLALLPPGAA-----STQVCTNNFTVSWFLRVKVSASHLEYOGCQVQ 57
 Db 14 LTLCAAGGALBEKKVCCQGSNRLTQGTGFEDHLSL-----QRYNNCEVVL 62
 Qy 58 GNLELTVLPTWASLFDQIEOVGYLIANNVOVPLORLIRIVRGTLQFEDNYALVL 117
 Db 63 GNLEITVQRYNDLSFLKTIQEVAGYLIANTVIRPLENQLIIRGNALYENTYALAIL 122
 Qy 118 DNGDPLNNTTPTVGTASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDI-----F 173
 Db 123 SN-----YGTNRTGLRELPMNLQELIGAVRFSNNPILCNMDTIQWRDIVQNVF 172
 Qy 174 HKNQLALTLIDNRSRACHPCSPMKGSRGRCWGSSEDCQSLRTVACGGCA-RCKGLPL 232
 Db 173 MSNMSMDL-----QSHPSCKPCDPCPNCGSCWGGEENCQKLTIIICAQQCSCRCGRSP 228
 Qy 233 TDCHEOCAAGCTGPKSHSDCLACLHFNHSGICELHCPALVTYNTDTTFESMNPGRVTFG 292
 Db 229 SDCCHNQCAAGCTGPRSDCLVCQKQFDEATCKTCTCPPLMYNTTVMQMDVNPBGKVSFG 288
 Qy 293 ASCVTACPNYLSLTDVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMEHLRE 352
 Db 289 ATCVKCKPRNVVTDHGSVCRAACGPDYVEV-EEDGIRCKKCDGFCRKCNGICIGEFKD 347

353 VRVTSANIQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPPQLQVFTLEBITGYLI 412
 Db 348 TLSINATNKHFKYCTAISGDHLHILVAFKGDSTFTPTPLDPRLEILKTVKEITGILLI 407
 Qy 413 SAWPDSLPLDSVFQNLQVIRGRILHNGAYSLTLOGLISWLGRLSRLSRELGGSLALIHHT 472
 Db 408 QAMPDNWTDLHAFENLEIIRGRTRKQHQGFSLVAVGLNITSLGRSLKEISDGDVILSNR 467
 Qy 473 HLCFVHTVPWQQLFRPHQALLHTANRPEDECVGEGLAHQLCARGHCWGPQTQVNCVS 532
 Db 468 NLCYANTINWKKLFGTPNQKTKIMNRAEKDCAVNVHVCNPLCSSBGCWGPFRDCVSCQ 527
 Qy 533 QFIRGQCEVEECVRLQGLPREYVNAHRLCPCHPCOPQNGSVTCFCPEADQCACAHYKD 592
 Db 528 NVSRGRCVCKNILEGEPREFVNESECIOCHEPCLPQANNITCTGRGPNCTQCAHYID 587
 Qy 593 PPFCAVRCPSGVKPDLSYMPINWKPDEBEGACQPCPINCTHSCVDLDDKDCAPARASPLT 652
 Db 588 GPHCVKTCPAGINGENNTL-VMKYADANNVCHLCHANCTYGCAGPGLQGCWVSPGPKIP 646
 Qy 653 SIYSANVGLILLVVVLGVVFGI-LIKRQKIRKVTMRRLLOETELVEPLTPSGAMPNQAQ 711
 Db 647 SIATGVGGLLFIW-VALGIGLPMRRHRTVRRKTLRLQLRELVEPLTPSGEAPNQAQ 705
 Qy 712 MRILKETELRKVKVLGSGAGFTVYKGIWIPDGNVPIPAIKVLRRENTSPKANKEILDEA 771
 Db 706 LRILKETEFKKIKVLGSGAGFTVYKGIWIPDGNVPIPAIKVLRRENTSPKANKEILDEA 765
 Qy 772 YVMAGVSPYVSRLLGICLTSTVQLTQMLPYGCLLDHVRNRRGLSGDQLNWCQIAK 831
 Db 766 YVMASVDNPHVCRLLGICLTSTVQLTQMLPYGCLLDYVREHKDNIGSYLLNWCQIAK 825
 Qy 832 GMSYLEVDRLVHRDLAARNVLKSPHVKITDFGLARLLDIDETEHADGCKVPIKWMAL 891
 Db 826 GMYLEDDRRLVHRDLAARNVLKTPQHVKITDFGLAKLLGAEKEYHAEGCKVPIKWMAL 885
 Qy 892 ESTLRRRFTHQSDVWSVGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDV 951
 Db 886 ESTLHRIYTHQSDVWSVGVTVWELMTFGSKPYDGIIPASDISILEKGERLPQPPICTIDV 945
 Qy 952 YIMVWKWIMIDSECRPRFRELVSFSEMRARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLE 1010
 Db 946 YIMVWKWIMIDASRPKRELILEFSQWARDPQRYLVIQGDERMHLPSPTSDSNFYRALMD 1005
 Qy 1011 DDMGDLVDABEYLVPOQGFCDPDPAGCGMVHRRHSSTSSRSGGDLTLGLEPSEEA 1070
 Db 1006 EEDMEDYVDADAYLIPQQGF-----NSPST----- 1031
 Qy 1071 PRSPLAPSEGAGSDVFDGLGMAAKGLQSLPTHDPSPLOYSDEPTVPLPSET--DGYV 1128
 Db 1032 SRPLLSLSLATS-----NSTVACINNGSCRVKEDAFLOYSDDPTGAVTENNIDDAFL 1087
 Qy 1129 APLTCSPPQPVYVQDPVRPQPPSPREGPLPAAPAGATLERAKTLPSPKNGVWVDVFAFG 1188
 Db 1088 -----PVPEYVQ-SVPKRPAGSVQNPVHNQPLHP-----APGRDLHYQN--PHS 1130
 Qy 1189 GAVENBEYL-TPQGAAPQPHPPAPSPAFDNLVYWDQ-----DP-----PERG 1231
 Db 1131 NAVGNPEYLNQA-----PTCLSSGFNSPALMTQKSHQMSLDNPDYQDDFFPKET 1181
 Qy 1232 APPSTFKGTPTAENPEYIGLDVP 1254
 Db 1182 KPNGIFKG-PTAENAEYLRVAPP 1203

RESULT 6
 TVCHLV
 epidermal growth factor receptor precursor - chicken
 N;Contents: protein-tyrosine kinase (EC 2.7.1.112) erbB
 C;Species: Gallus gallus (chicken)
 C;Date: 28-Feb-1986 #sequence_revision 05-May-1995 #text_change 04-Feb-2000
 C;Accession: A27720; A00643

R;Lax, I.; Johnson, A.; Howk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Vennstr
Mol. Cell. Biol. 8, 1970-1978, 1988
A;Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mou
A;Reference number: A27720; MUID:88261272; PMID:3260329
A;Accession: A27720
A;Molecule type: mRNA
A;Residues: 1-1223 <LAX>
A;Cross-references: GB:M20386
R;Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M
Cell 41, 719-726, 1985
A;Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A;Reference number: A00643; MUID:95228222; PMID:2988784
A;Accession: A00643
A;Molecule type: mRNA
A;Residues: 585-1223 <NTL>
A;Cross-references: GB:M10066
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
specific protein kinase
F;1-30/Domain: signal sequence #status predicted <SIG>
F;31-1223/Product: epidermal growth factor receptor #status predicted <MAT>
F;31-654/Domain: extracellular #status predicted <EXT>
F;81-307/Domain: EGF receptor extracellular domain repeat <BE1>
F;397-610/Domain: EGF receptor extracellular domain repeat <BE2>
F;655-677/Domain: transmembrane #status predicted <TM>
F;678-1223/Domain: intracellular #status predicted <INT>
F;719-984/Domain: protein kinase homology <KIN>
F;727-735/Region: protein kinase ATP-binding motif
F;136,202,280,361,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent) #
F;192,650/Binding site: carbohydrate (Ser) (covalent) #status predicted
F;687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F;754/Active site: Lys #status predicted
F;1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat

Query Match 45.38; Score 3086.5; DB 1; Length 1223;
Best Local Similarity 48.48; Pred. No. 7e-119;
Matches 628; Conservative 175; Mismatches 349; Indels 145; Gaps 25;

QY 8 RWGLLALLPPGAA-----STOVCTFNFTVSPFLVRPKVSAHSLLEYOGCQVQGNLE 61
DB 13 RGRAVLVLLLLGVALCSAVEKKVCCQGTNNKLTQLGHVEDHFTSLQRYNNCEVLNLE 72
QY 62 LTLPTNALSFLQDIQEVGYVLIHNRQVQLRQLRVGTQLPQEDNYALAVLDNGD 121
DB 73 ITVEHNRDLTFLKTIQEVAGVLIALNMVDVPLENLQIIRGNVLYDNSFALVLSNH 132
QY 122 PLNNTTPTVTCASPGGLRELQRLSLEILKGVLIQRNPQLCYQDTILWKDIFHNQNAL 181
DB 133 -MNKTO-----GLRELPMKRLSEILNGGVKISNNPKLCNMDTVLWNDIIDTSRK-PL 182
QY 182 TLID-TNRSRACHPCSPMKSGRCWGSSEDCSLTRTVCAAGCA-RCKGPLTDCHEQ 239
DB 183 TVLDFASSLSCPKCHPNCNTEHCHWAGEQNCQLTKVICAQCCSGRCRGKVPSCDCHNQ 242
QY 240 CAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTAC 299
DB 243 CAAGCTGPRESDCLACKFRDDATCKDTCPLVLVNYPTTYQMDVNPBGKYSFGATCVREC 302
QY 300 PNYLSTDVGSCTLVCPLNHNVQVTAEDGTORCSKSPCARVCYVGLGMEHLREVRVTS 359
DB 303 PHNYVVTDHGSCVRSNCNTDYEV-EENGVRKCKCDGLCSKVCNIGIGLKGILSINAT 361
QY 360 NIOEFAGCKKI FGLSLAPLPSFDGDPASNTAPLOPEQLQVFEITLGTGLYISAMPDSL 419
DB 362 NIDSFNCKTKINGDVSLLPVAFGLDAFTPLDLPKKLDVFRVKEISGFLLLQAMPDNA 421
QY 420 PDLVSFQNLQVIRGRILHNGAYSITLQGLGISWGLRLSLRELGLALIHNTHLCPVHT 479
DB 422 TDLVAFENLEIIRGTRKHQOQYSLAVNVNLIQSLGLRLSEISDGLIAMKNKVLVADT 481
QY 480 VPWDQLFRNPQALLHTANRPEDECVEGEGVGLACHQLCARGHCWGGPPTQCVCNCSQFLRGQE 539

RESULT 7

A47253
epidermal growth factor receptor, HER4 - human
C;Species: Homo sapiens (man)
C;Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C;Accession: A47253
R;Plozman, G.D.; Culoucou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.; N
Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993
A;Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epidermal
A;Reference number: A47253; MUID:93189574; PMID:8383326
A;Accession: A47253
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: nucleic acid
A;Residues: 1-1308 <PLO>
A;Cross-references: GB:I07868; NID:g337359; PIDN:AAB59446.1; PID:g337360
A;Note: sequence extracted from NCBI backbone (NCBIP:126842)
C;Superfamily: epidermal growth factor receptor; protein kinase homology

C; keywords: ATP; growth factor receptor
F; 716-981/Domain: protein kinase homology <KIN>
F; 724-732/Region: protein kinase ATP-binding motif

Query Match 43.6%; Score 2968.5; DB 2; Length 1308;
Best Local Similarity 45.2%; Pred. No. 4.8e-114;
Matches 612; Conservative 183; Mismatches 370; Indels 189; Gaps 30;

```
Qy 9 WGLLLALLPGAA-----STQVCTNNFTVSFWLVPKVSASHLB-----LYQGQV 56
Db 8 WVVSLVAAAGTVQPSDSQSVCAGTENKLS-----SLSDLEQQYRALRKYYENCEV 59
Qy 57 QGNLELYLPNNAISLFLODIOEQVGYVLLAHNOVROVPLQRLIRIVGTOLFDENYALAV 116
Db 60 MGNLEITSIEHNRDLSFRVRETVGVLVALNFRYLPLENLIIRGTIKLDYEDYALAI 119
Qy 117 LDNGDPLNNTPPVGTASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIPHN 176
Db 120 FLNRYKDGNF-----GLQELGLKLNTEILNGVGVVQNKFLCYADTIHQDVRNP 170
Qy 177 NQLALTILIDNRSRACHPCSPMGKSGRCWGBESSDCSLTRTVACAGC-ARCKGPLTDC 235
Db 171 WPSNLLTLVSTNGSSGCRCHKSTG-RCWGPTEHNCQTLTRTVCAEQDCRCYGPVYSDC 229
Qy 236 CHEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTESMPNPEGRVTFGASC 295
Db 230 CHREACGCGSPKDTDFACFNFDNSGACVTQCQTFVYNTPTTQLEHNFNAKITYGAF 289
Qy 296 VTACPYNLSTDVGSCTLVCPHLNHQVTAEDGTORCEKCKPKCARVCYGLGMEHLREVR 355
Db 290 VKKCPHNFV-VDSSSVCACRACSPSSRMVEV-EENGIRKCKPCTDICKPKADGIGTGLMSAQT 347
Qy 356 VTSANIOEPACCKIPGSLAPLPSFDGDPASNTAPLOPEQLQVFTLEBITGLYLIYISAW 415
Db 348 VDSNIDKFNCTKINGNLIFLVTGIHGDPPYNAIEADPEKLNVRFTVETGFLNQS 407
Qy 416 PDSLPDLVSFONLQVIRGRIILNCAYSITLQGLGSIWGLRSRLSGLALIHNNHLC 475
Db 408 PPNMTDFVSFNLTIGRVLISGLSLILILKQGITSLQFQSLKEISAGNIYITDNLCL 467
Qy 476 FVHTVPMDQLFRPNHQALLTANRPEDECVGEGACHQLCARGHCWGPPTQCVNCSQFL 535
Db 468 YHTINNTLFTSTINQIRIVIRDNKKAENCTAEGVMCNHLSCSSDCWCGPDPQCLSCRRFS 527
Qy 536 RGQECVCECVRLQGLPREYNARCLCHPECPQ-QNGSVTCFQPEADQVACAHYKDP 594
Db 528 RGRICIESNLYDGEFREFENGSIQVCECDQCEKEDGLTCHGPGPDNCTKSHFKDGP 587
Qy 595 FCVARCPGVKPDLSYMPIWKFPEEGACQPCINCTHSCVDLDDKGC-----P 643
Db 588 NCVKCPDGLQGANSF--IFKADPDRECHPCNCTQCGNGPTSHDCIYYPWTGHSTLP 645
Qy 644 AEQRASPLTSIVSAV-GILLVVLGVVGFILIKRROQKIRKTYMRLLQETELVEPLTP 702
Db 646 QHAR-TPL-IAAGVIGLFLIVGLTFAVYVRKSIK-KKRALRRFL-ETELVEPLTP 700
Qy 703 SCAMPNOAQRILKTELKRVKVLGSAFGVYIGWIPDGENKIPVAIKVLRENTSPK 762
Db 701 SGTAPNOAQLRKTELKRVKVLGSAFGVYIGWIPDGETVKIPVAIKILNETTGP 760
Qy 763 ANKEILDEAYVWAGVSPYRLLGICLTSTVQLVTOLMPYGCLLDHVRENRLGSDQL 822
Db 761 ANVEFMEALIMASDHPIVRLILGVCLSPILQVLTQMPHGCLLEYVHEKONIGSQL 820
Qy 823 LNCWQIAKMSYLEVRLVRLDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGG 882
Db 821 LNCWQIAKGMYLEERLVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGG 880
Qy 883 KVPILKWALESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGTIPAREIDPLEKGERLP 942
Db 881 KNPILKWALECIHYRKETHQSDVMSYGVTVWELMTFGKPYDGTIPAREIDPLEKGERLP 940
Qy 943 QPPICTIDVYIMVKCMIDSECRPRELVSSEFSRMARDPQRFVWIQNED-LGPASPLD 1001
```

```
Db 941 QPPICTIDVYIMVKCMIDADSRPKFKELAAEFSSMARDPQRYLYIQGDRMKLPSND 1000
Qy 1002 STFYRSLLEDDMGDLVDAEYLVPOQGFCCPDPAFCAGGVHRRSSSTRSGGDLTL 1061
Db 1001 SKFFQNLLDEEDLEDMDAEYLVLP-QAFNIPPP-----IYTSRARIIDNSR-----EI 1048
Qy 1062 GLEPSEEEAPRS-----PLAP-SEGAGSDVFDGDLGM 1092
Db 1049 GHSPPPAYTPMSGNQFYVRDGGFAAQGVSVYRAPTSITPEAPVAGATAEIFDDSCCN 1108
Qy 1093 GAAKGLQLSPLTHDPSPQLQRYSEDTVPLPS-----ETDGYVAPLTCSQPQYVNPQDV 1145
Db 1109 GTLRKFPVAHVQDSSTQRYSDPTVFAPERSPRGELDEEGYTMPRDKPKQEVLNPE- 1167
Qy 1146 RPOPPSPRGPLPAAPAGATLERAKTSLSPGKGVKVDVAFGGAIVENPYLTPQGGAA 1205
Db 1168 -----ENPFVSR-----KNGDLQ-----ALDNPYHNASG--- 1194
Qy 1206 QPHPPPA-----FSPAFDNLVYWDODPPERG 1231
Db 1195 ---PPKAEDEYVNEPLYLNTFANTLGAEYLNILSMPEKAKAFDNDPDYWNHSLPPRS 1251
Qy 1232 A--PPSTFKGTP-----AENPEYL 1249
Db 1252 TLQHPDYLQEYSTKYFKQNGRIRPIVAENPEYL 1285
RESULT 8
S06142
protein-tyrosine kinase (EC 2.7.1.112) mrk-y precursor - southern platyfish
N; Alternate names: epidermal growth factor receptor homolog; kinase-related transformin
C; Species: Xiphophorus maculatus (southern platyfish)
C; Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 18-Feb-2000
C; Accession: S06142; S13809
R; Mitterbrodt, J.; Adam, D.; Malitschek, B.; Maeueller, W.; Raulf, F.; Telling, A.; Robert
Nature 341, 415-421, 1989
A; Title: Novel putative receptor tyrosine kinase encoded by the melanoma-inducing Tu lo
A; Reference number: S06142; MUID:90015140; PMID:2797166
A; Accession: S06142
A; Molecule type: DNA
A; Residues: 1-1166 <WIT>
A; Cross-references: EMBL:X16891; NID:G65290; PIDN:CAA34770.1; PID:G65291
R; Adam, D.; Maeueller, W.; Schartl, M.
Oncogene 6, 73-80, 1991
A; Title: transcriptional activation of the melanoma inducing Xmrk oncogene in Xiphophor
A; Reference number: S13807; MUID:91125882; PMID:1846957
A; Accession: S13809
A; Status: preliminary; translation not shown
A; Molecule type: DNA
A; Residues: 821-1025, 'N', 1027-1098, 'A', 1100-1166 <ADA>
A; Cross-references: EMBL:X56319; NID:G65284; PIDN:CAA39763.1; PID:G65285
C; Genetics:
A; Gene: mrk
A; Map position: Y
A; Introns: 872/3; 898/1; 947/1; 979/3; 1025/3; 1056/1
C; Superfamily: epidermal growth factor receptor; protein kinase homology
C; Keywords: ATP; growth factor receptor; phosphotransferase; transmembrane protein; tyr
F; 1-25/Domain: signal sequence #status predicted <SIG>
F; 26-1166/Product: kinase-related transforming protein (Tu) #status predicted <MAT>
F; 707-972/Domain: protein kinase homology <KIN>
F; 715-723/Region: protein kinase ATP-binding motif
Query Match 39.5%; Score 2689; DB 1; Length 1166;
Best Local Similarity 46.1%; Pred. No. 1.1e-102;
Matches 562; Conservative 156; Mismatches 372; Indels 129; Gaps 25;
Qy 48 ELYQGVQVQGNLELTVLPTNASISFLDIOEQVGYVLLAHNOVROVPLQRLIRVGTQL 107
Db 51 KMTSGCNVNVLENLEITITQENQDLSLQSIQEVGGVLLIANNEVSTIPLVNLRLIRQNL 110
Qy 108 FENYALAVLDNGDPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTI 167
```

Db 111 YEGNFTLLVMSNYOK- NPSSP- -DVYQVGLKQLQSLNTEILSGGVKVSHPNLLCNVETI 167
Qy 168 LMKDIFPHKNNQALTLITDNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVCAAGC-AR 226
Db 168 NWWDIVDKTSNPTNWLIPHAFCQKQCDHCVNGSCWAPGPGHCQKFTKLLCAEQCNRR 227
Qy 227 CKGPLPTDCCHQEOCAAGCTGPKHSDCLACLFHNSHGICELHCPALVTYNTDTFESMNPDE 286
Db 228 CRGPKPTDCCNHCAAGCTGPRATDCLACRDNFDGTCDFCPKPKYDVIHSHQVNDPN 287
Qy 287 GRYTFGASCVTACPVYVLTSDVSGTLCVPLHNEVTAEDGTORCEKSPCARVCYGLG 346
Db 288 IKYTFGACVKECSNVVTE- GACVRSAGMLEVD- ENKRSCKPCDGVCPKCDGIG 345
Qy 347 MEHLREVRVTSANIQFAGCKIFGLAFLPESFDGDPASNTAPLOPEQLQVETILEE 406
Db 346 IGSLSNTIAVNSTIRSFNSCKINGDIIILNRNSFEGDPHYKIGTMDPEHLNLTVKEI 405
Qy 407 TGYLYISAWPDSLPDLVSFQNLVIRGRILHNGAYS- LTLQGLIGISWGLRSLRELSG 465
Db 406 TGYLVIMWPNMTSLVSFQNLVIRGRITFTSRGFSFVVQVRHLQWGLRSLKEVSAGN 465
Qy 466 ALIHNTLHLCFVHTVPMDLFRNPHOALLHTANRPEDECVGEGLAGHOLCARGHCWGP 525
Db 466 VILKNTIQLRYANTINRRILFRSDQSIYDART- - - - - ENTCNNECEDGCW- PGP 517
Qy 526 TQVCNCSQFLRGQCEBCEVRLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCV 585
Db 518 TMCVSCLVHVRGRCVASCNLLQGEPREAQVDGRCVQCHOECLVOTDSLTCYGPANCS 577
Qy 586 ACAHYKDPFCVACPCSGVPKDSYMPIWKPDBEGACQCPINCTHSCVDLDDKGPAP 645
Db 578 KSAHFODGPOQIPRCPHGILGDGDTL- IWKYADKMGQCQCHQNTQCGSGPGLSGCRGD 636
Qy 646 QRASPLTSIYSAVGIILVVLVGVFGLIKRROOKIRKYTMRELLQETLVEPLTSGA 705
Db 637 -IVSHSLAVGLVSGLLITVIVALLIVLLRRRIK- RKRITICLLQEKELVEPLTSGQ 694
Qy 706 MPNQAMRIILKETELRKVKVLGSGAFGVYKGIWIPDENVKIPVAIKVLRNTPSKANK 765
Db 695 APNQAFRIILKETEFKDRVLGSGAFGVYKGLWNPGENIRIPVAIKVLRNTPSKANK 754
Qy 766 EILDEAVMAGVSGPYVSRLLIGICLTSTVQLVTOLMPYGCLLDHVRENRLGSGDILLN 825
Db 755 EVLDEAVMASVDHPVHCRLLIGICLTSAVQLVTOLMPYGCLLDVYRQHERICQWLLN 814
Qy 826 CMQIAKMSYLEDLVRLHRLDAARNVLKSNHVKITDFGLARLLDIDETEHADGKVP 885
Db 815 CVQIAKMNYLEERHLVHRLDAARNVLLKNPNHVKITDFGLSKLLTADKEXYQADGKVP 874
Qy 886 IKWMALESILRRRPTHQSDVMSYGVTVWELMTFCAKPYDGIPIAREIPOLLEKGERLP 945
Db 875 IKWMALESILQWYTHQSDVMSYGVTVWELMTFQSKPYDGIPIAREIASVLENGERLP 934
Qy 946 ICTIDVIMVKMIDSECRPRELIVSEFSRWARDPQRVYQNEEDLGPASPLDSTFY 1005
Db 935 ICTIEVTMILKCMIDPSSRPRELIVSEFSWARDPFRVLIQ- - - NLPSSLDRLF 991
Qy 1006 RSLLEDDMDGLVDAEYLPQGGFFCDPAPGAGGVHHRSSSTRSGGDLTLGLEP 1065
Db 992 SRLSSDD- - - DVVDAEYLLPYKRI- - - - - NROGS- - - - - 1019
Qy 1066 SEEAAPRSPLAPSEAGSDVFDGLGMAAKGLQSLPHTDPSPLQRYSEDPV- PLPSET 1124
Db 1020 - - - - - EPCIPPTGH- - - - - PVRENSITLRNISDPTQNALEKDL 1052
Qy 1125 DGYVAPLTCFQPPYVNPQDVRPQ- - - - - PSPRE- - - - - GLP- AARPAGATLBR 1171
Db 1053 DGH- - - - - EYVNPQGETSSRLSDIYNPNVEDLTDGWPVPSLSQEAETNFSRPE 1102
Qy 1172 TLPCKNGVVKVDFAGGAVENPEYLTPOGGAAPQHPPPAFNLYWDDQPPERG 1231
Db 1103 YLANTNQSL- - - - - PLVSSGSMDDPDY- - - - - YQAAF- - - - - LPQTG 1137

Qy 1232 APPSTFKGTPTAENPEYLG 1250
Db 1138 ALTGNGMFLPAAENLEYLG 1156

RESULT 9

A36223
C:Species: Homo sapiens (man)
C:Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000
C:Accession: A36223; I59164
R:Kraus, M.H.; Iseling, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.
Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989
A:Title: Isolation and characterization of ERB3, a third member of the ERBB/epidermal
A:Reference number: A36223; MUID:90083234; PMID:2687875
A:Accession: A36223
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1342 <KRA>
A:Cross-references: GB:M29366
R:Plowman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todaro, G.
Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990
A:Title: Molecular cloning and expression of another epidermal growth factor receptor-r
A:Reference number: I59164; MUID:90311312; PMID:2164210
A:Accession: I59164
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-559, 'G', 561-957, 'F', 959-1063, 'G', 1065-1342 <RES>
A:Cross-references: GB:M34309; NID:G183990; PIDN:AAA35979.1; PID:G306841
C:Genetics:
A:Gene: GDB:ERB3; HER3
A:Cross-references: GDE:119880; OMIM:190151
A:Map position: 12q13-12q13
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homol
C:Keywords: ATP; phosphotransferase
F:707-972/Domain: protein kinase homology <KIN>
F:715-723/Region: protein kinase ATP-binding motif

Query Match 35.1%; Score 2408; DB 2; Length 1342;
Best Local Similarity 40.3%; Pred. No. 3.6e-91;
Matches 530; Conservative 191; Mismatches 447; Indels 146; Gaps 33;
Qy 5 ALCRWGLLLALLPPGAASTQVCTFNFTVSWLRVFKVSASHLEYQCGQVQGNLELY 64
Db 27 AVCP-GLNGLSVTGDENQYQTL- - - - - YKLYERCEVWGNLEIVL 67
Qy 65 LPTNASLSFLQDIOEVQGVYVLIHNOVROVPLORLIRVGTQLFEDNYALVLDNGDPLN 124
Db 68 TGNADLSFLQVIREVTGYVVMANBFSTLPLNLRVVRGTQVYDGKFAIFVM- - - - - LN 122
Qy 125 NTPVTGASPGGLRELQLRSLTEILKGGVLIQBNPOLCYQDITLWKDI PHKNQLALTLI 184
Db 123 YNT- - - - - NSSHAURLQLTQLTEILSGGVYIEKNDKLCMDTIDWRDIVDRD- - - - - AEIV 175
Qy 185 DTVNRACHPCSPMKSGRCWGSSEDCOSLRTVCAGGC- ARCKGPLPTDCCHQEOCAAG 243
Db 176 VKDNGRSCFPCHEVCKG- RCWGPGESEDCQTLTKTI CAPQCNHCFGFPNQCCHDECAGG 234
Qy 244 CTGPKHSDCLACLFHNSHGICELHCPALVTYNTDTFESMNPPEGRYTFGASCVTACPYNY 303
Db 235 CSGPQDTCFACRHFNDGACVPRCPQPLVYNKLTQLQEPNPHTKYQYGVCVASCPSHNF 294
Qy 304 LSTDVSGCTLVCPHMQEVTAEADGTORCEKSPCARVCYGLMEHLRVRVATYSANIQE 363
Db 295 V-VDQTSVCRACPPDKMEVD- - - - - KNGLKMCPCGGLCPKACEGTGSG- - - - - SRFTYDSSNIDG 350
Qy 364 FAGCKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFETLEETGYLYISAWPDSLPDL 423
Db 351 FVNCTKILGNLDFLITGLNGDPWHKIPALDPEKLNIVFRVREITGYLNTQSWPHHNF 410
Qy 424 VFQNLQVIRGRILHNGAYS- LTLQGLIGISWGLRSLRELGSGLALIHNTLHLCFVHTVP 482

Db 411 VFSNLTITIGRSLYNRGFSLLIMKNLNVTSIGPRSLKEISAGRIYISANRQLCYHHSLNW 470
Qy 483 DQLFRNPHQALLHTA-NRPEDECVGEGLAHQLCARGHCWGPQTQCWNCQSFRLRGQECV 541
Db 471 TKVLRGTEERLDIKHNRPRDCVAEGKVCPLCGSSGCGWGPQGCLSCRNYSGRGVCV 530
Qy 542 EECRVQLGLPREYVNAHCLPCHPECOPOGNSVTCFGEADOCVACAHYKDPFPCVACRCP 601
Db 531 THCNFLNGEPREFAHEACFSCHEPCQDMEGTATCNGSGSDTCAQCAHFRDGPCHVCSSCP 590
Qy 602 SGVKPDLISYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCAPBQRA-----SPLTSIVSA 657
Db 591 HGVLG--AKGPIYKVPQVQNECRCHENCCTQCKGPELQDCLGQTLVLIGKTHLTMTALT 648
Qy 658 VVGLLVVVLGVVFGIILIKRQOKIR-KYTMRRLLQETELVEPLTPSGAMPNQOMRLK 716
Db 649 IAG--LVVIFMMLGGTFLYWRGRIQNRAMRRYLERGESIEPLDPS-EKANKVLARIFK 705
Qy 717 ETELKRVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILD EAYVMAG 776
Db 706 ETELKRVKVLGSGVFTYKGVWIPDGENVKIPVAIKVLRNTSPKANKEILD EAYVMAG 765
Qy 777 VGSYVSRLLGICLTSTVQLVTLQMPYXGCLLDHVRNRLGRSGDILLNMCQIAKMSYL 836
Db 766 LDHAHIVRLGLCPGSSQLVTVQLPLGSLLDHVRQHRGALGPQLLLNWGVQIAKGMYYL 825
Qy 837 EDVLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMMALSLR 896
Db 826 BEHGMVHRNLAAARNVLKSPQVQVADFGVADLLPPDDKQLYSEAKTPIKMMALSIHF 885
Qy 897 RRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYMW 956
Db 886 KXYTHQSDVMSYGVTVWELMTFGAEYAGLAEVDPDLLEKGERLAQOICTIDVYMWV 945
Qy 957 KCMWIDSECRPRFRELSEFSEFARMARDPQRFVVIQNEIDLGA-----SPLDSTFYRSLLEDD 1013
Db 946 KCMWIDENIRPTKELANEFTRMARDPPYLVIRKES-GPGIAPGPRPHGLTNKKLEVE 1004
Qy 1014 MGLVDABEYLVPOGFCPPDAPGAGGVWHRHRSSTRSGGDLTLIGLEP-SEEAAPR 1072
Db 1005 LEPELDLDLLEAED-----NLATTGLSALSLPVGTLNRRGSG 1045
Qy 1073 SPLAPSEGAGSDVFDGDLGMAAGLOSPLTHD-PSPLORYSEDTVPPLP-----SETD 1125
Db 1046 SLLSPSSGY-MPMNQNLGESCQSAVSGSSERCPVSLH-----PMRGCLASESE 1098
Qy 1126 GYVA-----PLTCSPOPE-----YVNPQDVPRPQPPSPREGP----- 1156
Db 1099 GHVTGSEAEELQEKVSMCRSRSRSPRPGDSAYHSQRHSLLTPTVPLSPGLEEDVNG 1158
Qy 1157 --LPAARPAGATLERAKTLSP-KNGVV-----KDVFAFGAVENPEYLTPOGGAPOP 1207
Db 1159 YVMPDTHLKGTPSSREGTLSSVGLSGLTGEED-----EYEVNRRRRHSP-P 1209
Qy 1208 HPPPAFSPAFNLVYWD-----QDPPERGAPPSTFKGTPTAENPEYL 1249
Db 1210 HPPRPSLEELGYEYMDVGSLSASLSTQSCPLHPVIMPETAGTTPDEDETEYM 1263

RESULT 10

JC4387
epidermal growth factor receptor homolog precursor - rat
N:Alternate names: ErbB3 protein; HER3 protein
C:Species: Rattus norvegicus (Norway rat)
C>Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998
C:Accession: JC4387
R:Hellver, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.
Gene 165, 279-284, 1995
A>Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protein.
A:Reference number: JC4387; MUID:96096535; PMID:8522190
A:Accession: JC4387
A:Molecule type: mRNA
A:Residues: 1-1339 <HEL>

A;Cross-references: GB:U29339; NID:G915389; PID:G915390
A;Experimental source: liver
A;Note: The authors translated the codon AAC for residue 369 as Thr and GTT for residue 370 as Lys. This protein is a functional heregulin receptor that transduces signals to C;Genetics:
A;Gene: ErbB3
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homol. C;Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane protein F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1339/Product: epidermal growth factor homolog #status predicted <MAT>
F;640-659/Domain: transmembrane #status predicted <TM>
F;705-970/Domain: protein kinase homology <KIN>
F;713-721/Region: protein kinase ATP-binding motif
F;939,1051,1156,1194,1196,1219,1257,1273,1286,1325/Binding site: phosphate (Tyr)

Query Match 34.2%; Score 2327.5; DB 2; Length 1339;
Best Local Similarity 40.6%; Pred. No. 6.9e-88;
Matches 519; Conservative 170; Mismatches 422; Indels 167; Gaps 34;

Qy 5 ALCRWGLLLALLPPGAASQVCTFNFTVSFVLRVVKVSASHLELYCGCOVQGNLELY 64
Db 27 AVCP-GTLNGLSVTGADNQYQL-----YKLYKECEVVMGNLEIVL 67
Qy 65 LPTNASLSFLQDIQEVQGVYLIAHNOVRQVPLQRLRIVRGTLQFEDNYALAVLDNGDPLN 124
Db 68 TGHNADLSFLQIREVTAYVVLVAMNEFSLPLNLRVVRGTQVYDQKFAIFM----LN 122
Qy 125 NTPVTGASPGGURELQRLSLTEILKGVLIQRNPOLCYQDTILKWDIFHNNQLAULTI 184
Db 123 YNT-----NSHALRQLKFTQLTEILSGGVYIERDKLCHMDTIDWRDIVR---GAEIV 175
Qy 185 DTVRSRACHPCSPMKGSRGWSSESSDQSLTETVCAGGC-ARCKGPLPTDCHEOCAA 243
Db 176 VRNNGANCPCPEVCKG-RCWGFDPDCLLTITTCAPQNGRCFGFNPQCCHDECAGG 234
Qy 244 CTGPKHSIDCLALHFNHSGICELHCPALVTYNTDTFESMPNPGRTYFGASCVTACPYNY 303
Db 235 CSGPQDTDFACRRFNDGACVPCPEPLVYNKLTQLQENPHTKYQYGGCVASCPCPHF 294
Qy 304 LSTVDSGCTLVCPHINQEVTAEDGTORCEKSPCARVCYGLGMEHLREVRATVSANIQE 363
Db 295 V-VQDTFCVRACPPDKMEVD-KHGLKMCPECGGLCPKACEGTSG--SRYQTVDSSNIDG 350
Qy 364 FACKKIFGSLAFLPESFDGDPASNTAPLOEQLQVFELEITGVLYISAWPDSLPLDS 423
Db 351 FVNTKILGMLDLITGLNVDPWHKIPALDPEKLVNFRVREITGVLYINISQSWPHMNF 410
Qy 424 VFQNLQVIRGRIILHNGAYS-LTLQGLGISWGLSLRSLRSLGSLALIHNNHLFCVHTVPW 482
Db 411 VFSNLTITIGRSLYNRGFSLLIMKNLNVTSIGPRSLKEISAGRIYISANRQLCYHHSLNW 470
Qy 483 DQLFRNPHQALLHTA-NRPEDECVGEGLAHQLCARGHCWGPQTQCWNCQSFRLRGQECV 541
Db 471 TKVLRGTEERLDIKHNRPRDCVAEGKVCPLCGSSGCGWGPQGCLSCRNYSGRGVCV 530
Qy 542 EECRVQLGLPREYVNAHCLPCHPECOPOGNSVTCFGEADOCVACAHYKDPFPCVACRCP 601
Db 531 THCNFLNGEPREFAHEACFSCHEPCQDMEGTATCNGSGSDTCAQCAHFRDGPCHVCSSCP 590
Qy 602 SGVKPDLISYMPIWKFPDEEGACQPCINCTHSC-----VDLDDKGCAPAEQASPLTSIVSAV 659
Db 591 HGVLG--AKGPIYKVPQVQNECRCHENCCTQCKGPELQDCLGQAEVLMKSPHLVIAVTV 648
Qy 660 GILLVVVLGVVFGIILIKRQOKIR-KYTMRRLLQETELVEPLTPSGAMPNQOMRLKET 716
Db 649 G--LAVITMLTGGSLYWRGRIQNRAMRRYLERGESIEPLDPS-EKANKVLARIFKET 705
Qy 719 ELRKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGV 778
Db 706 ELRKVLKVLGSGVFTYKGVWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGV 765
Qy 779 SPYVSRLLGICLTSTVQLVTLQMPYXGCLLDHVRNRLGRSGDILLNMCQIAKMSYLE 838
Db 766 LDHAHIVRLGLCPGSSQLVTVQLPLGSLLDHVRQHRGALGPQLLLNWGVQIAKGMYYL 825

Db 715 QYTAIGY-----CAASPPRSSKITANLD-----YNNIFITGAVLVPTIC 755
Qy 669 VVFGI-LIKRQOKIRYK--MRLLOQETELVEPLTPSGAMPNQAQMRILKETELRKVKV 725
Db 756 ILCVVTVICROKQAKKETVQMTWALSQREDSFRLPSNIGANLCKLRIVKDAELRGGV 815
Qy 726 LGSAGFTVYKGIWIPDGENVKIPVAIKVRENTSPKANKEILDEAYVMAGVGSFYVSR 785
Db 816 LGMAGFRVYKGVWVPGENVKIPVAIKELLKSTGABSSEFFLEAYIMASEEHVNLKL 875
Qy 786 LGICLTSTVOLATOLMPYVGLDHRNRCGLSGODLLNMCQIAKMSYLEDLVRLVHRD 845
Db 876 LAVCMSSQMLITOLMPLGCLLDIVRNRRDKIGSKALLNMQSTQAKMSYLEEKLRLVHRD 935
Qy 846 LAARNVLVK---SPNHVKITDFGLARLLDIDETEHADGGKVPKMKWMALESILRRRTHQ 902
Db 936 LAARNVLRLLAGEDH---DFGLAKLLSSDSNEYKAAAGKMPKMLALCIRNRVTSK 991
Qy 903 SDVMSYGVVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPCTIDVYMWIMVKWMD 962
Db 992 SDVMAFGVTIWEILLTFQORPHENIPAKDIPDLIEVGLKLEQETCSLDIYCTLLSCWHL 1051
Qy 963 SECPRELVSERSMARDPQRFVWQNEGLG--PASPLDSTFYRSLLEDD---DMGDL 1017
Db 1052 AAMRPTFKQITTVFAEFARDPGRYLAITLQDKFTRLPA-----YTSODEKOLIRKLAPT 1104
Qy 1018 VDABEYLVPOQGFPPAPGAGGMVHRRSSSTRSGGDLTGLGPESEEAAP----- 1071
Db 1105 TDGSEATKDDYLOPKAALGPS-----HRTDCT-----DEMPKLNRYC 1143
Qy 1072 RSLAPSEAGSDVFDG---DLGMAAGKGLQSLTHDPSLQRYSEDTVPLPSETDGYV 1128
Db 1144 KPSNKNSSGDDERDSAREVGVNLR-----LDLPVDEDDYL 1182
Qy 1129 APLTCSQPEYVQPDVPPSPREGPLPAARPAAGATLERAKTLSPGKGVVKDPAFG 1188
Db 1183 MP-TCQPGPNNNMN-----NPNQNNMAAVGAAGYM-----DLIGVP 1220
Qy 1189 GAVENPEYL---TPQGAAPQPH-----PPAFSP-AFDNLYYWD 1224
Db 1221 VSDVNPYLLNAQTLGVGSEPIPTQITIGIPWGGPGTMEVKVPMFGSEPTSSDHEYYND 1279

RESULT 14

S35745
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus
C;Species: avian erythroblastosis virus
C;Date: 03-Mar-1994 #sequence_revision 26-May-1995 #text_change 28-Feb-1997
C;Accession: S35745
submitted to the EMBL Data Library, March 1993
A;Reference number: S35743
A;Accession: S35745
A;Molecule type: DNA
A;Residues: 1-544 <VEN>
A;Cross-references: EMBL:X12707
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific p
F;135-400/Domain: protein kinase homology <KIN>
F;143-151/Region: protein kinase ATP-binding motif
F;170/Active site: Lys #status predicted

Query Match 24.2%; Score 1647; DB 2; Length 544;
Best Local Similarity 54.9%; Pred. No. 1.6e-60;
Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15;

Qy 578 GPEADQCVACAHYKDPFPCVARGCPSPKPDLSYMPIWKFDEBAGCQPCINCTHSCVDL 637
Db 1 GP---DHCMKCAHFIDGPHCVKACPAAGVLGENDTL-VWKYADANAVCOLCHPNCNTRGCKGP 57

Qy 638 DDKGCPAEGRASPLTSIVSAV--GILLVVVLGVVFGILIKRROOKIRKYTMRRLLQETEL 696
Db 58 GLEGCP---NGSKTPSIAAGWGGLCLVAVVGLIGLYLRRR-HIVKRTLRLLQEREL 113
Qy 697 VEPITSGAMPNQAQMRILKETELRKVKVLSGAFGTVYKGIWIPDGENVKIPVAIKVL 756
Db 114 VEPITSGEAPNQAHRILKETELFKVKVLSGAFGTVYKGLWIPGEKVTIPVAIKELR 173
Qy 757 ENTSPKANKEILDEAYVMAGVGSFYVSRLLGICLTSTVOLVTLQMPYVGLDHRNRCGR 816
Db 174 EATSPKANKEILDEAYVMAVDNPHVCRLLGICLTSTVQLITOLMPYVGLDLYREHKDN 233
Qy 817 LGSQDLLNMCQIAKMSYLEDLVRLVHRDLAARNVLKSNHVKITDFGLARLLDIDETE 876
Db 234 IGSQYLLNMCVQIAKGMNYLEERHVMVHRDLAARNVLKTPQHVKITDFGLAKQLGADEKE 293
Qy 877 YHADGGKVPKMKWMALESILRRRTHQSDVMSYGVVWELMTFGAKPYDGIPIAREIPDLLE 936
Db 294 YHAEKGKVPKMKWMALESILHRIYTHQSDVMSYGVVWELMTFGSKPYDGIPIAREISSVLE 353
Qy 937 KGERLPQPPCTIDVYMWIMVKWMDSECRPRPRELVSERSMARDPQRFVWQ-NEIDL 995
Db 354 KGERLPQPPCTIDVYMWIMVKWMDSDADSPKRELIAEFSKWARDPPRVLYIQGDRMH 413
Qy 996 PASPLDSTFYRSLLEDDMGDLVDABEYLVPOQGFPPAPGAGGMVHRRSSSTRSG 1055
Db 414 LPSPTDSKF/RTLMEBEDMEDI VDAEYLVPHQGF-----NSPST--- 454
Qy 1056 GGLTLGLSESEEAAPRSL-----APSEGAGSDVFDGDLGMAAGKGLQSLTHDPSPLQ 1110
Db 455 -----SRTPLLSLSATSNNSATNCIDRNGG-----H----- 481
Qy 1111 RYSEDPVPLPSETDGYVAPLTCSPQPEYVQPDVPPSPREGPLPAARPAAGAT-LER 1169
Db 482 -----PVREDGFL-----PAPEYVQ--LMPKPESTAMVQNIYVLSLTISK 523
Qy 1170 AKTLSPGKGVVKDVPFAFGGAVENPEYL 1197
Db 524 LPIDRSYQN-----SHSTAVDNPEYL 544

RESULT 15

S00727
kinase-related transforming protein (erbB) (EC 2.7.1.1.-) - avian erythroblastosis virus
C;Species: avian erythroblastosis virus
C;Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 28-Feb-1997
C;Accession: S00727
R;Scotting, P.; Vennstrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.
Oncogene Res. 1, 265-278, 1987
A;Title: Common site of mutation in the erbB gene of avian erythroblastosis virus mutan
A;Reference number: S00727; MUID:88217326; PMID:2897102
A;Accession: S00727
A;Molecule type: DNA
A;Residues: 1-545 <SCO>
A;Cross-references: EMBL:X06943
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; phosphotransferase
F;135-400/Domain: protein kinase homology <KIN>
F;143-151/Region: protein kinase ATP-binding motif

Query Match 24.1%; Score 1640; DB 2; Length 545;
Best Local Similarity 54.9%; Pred. No. 3.1e-60;
Matches 345; Conservative 69; Mismatches 122; Indels 92; Gaps 15;

Qy 578 GPEADQCVACAHYKDPFPCVARGCPSPKPDLSYMPIWKFDEBAGCQPCINCTHSCVDL 637

Db 1 GP---DHCMKCAHFIDGPHCVKACPAAGVLGENDTL-VWKYADANAVCOLCHPNCNTRGCKGP 57

Qy 638 DDKGCPAEGRASPLTSIVSAV--GILLVVVLGVVFGILIKRROOKIRKYTMRRLLQETEL 696
Db 58 GLEGCP---NGSKTPSIAAGWGGLCLVAVVGLIGLYLRRR-HIVKRTLRLLQEREL 113

```
Qy 697 VEPLTPSGAMPNOQMRLKETELRKVKVILGSGAGFTVYKGIWIPDGENVKIPVAIKVLR 756
Db 114 VEPLTPSGEAPNOAHLRLKETEFKKVKVILGFGAGFTVYKGLWIPEGEKVTIPVAIKELR 173
Qy 757 ENTSPKANKEILD EAYVYVAGVSPVSRLLGICLTSTVQLVTLMPYGCCLLDHVRENRR 816
Db 174 EATSPKANKEILD EAYVYVWASVDNPHVCRLLGICLTSTVQLITQLMPYGCCLLDYIREHKDN 233
Qy 817 LGSODLLNWCQIAKGSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETE 876
Db 234 IGSQYLLNWCQIAKGMNYLEERHLVHRDLAARNVLKTPQDVKITDFGLAKQLGADEKE 293
Qy 877 YHADGKVPISKMALESTLRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLLE 936
Db 294 YHAEKKVPISKMALESTLHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPIASEISSVLE 353
Qy 937 KGERLPQPPICTIDVYIMVWCWIMIDSECRPRFRELVSFESRMARDPQRFVVIQ-NEDLG 995
Db 354 KGERLPQPPICTIDVYIMVWCWMSDADSRPKFRELIAEFKSMARDPPRYLVIQDERMH 413
Qy 996 PASPLDSTFYRSILLEDDMGDLVDAEYLVPOQGFPCDDPAPGAGGMVHHRRSSSTRSG 1055
Db 414 LPSPTDSKFYRTLMEEDMEDIVDAEYLVPHQGF-----NSPST--- 454
Qy 1056 GGDLTGLPSEEEAPRSP-----APSEGAGSDVFDGDLGMAAKGLQSLPTHDPSPQLQ 1110
Db 455 -----SRTPLLSLSATSNNSATNCIDRNG-----H----- 481
Qy 1111 RYSEDPTVPLPSETDGVYVAPLTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAGAT-LER 1169
Db 482 -----PVREDGFL-----PAPEYVNQ--LMPKKPSTAMVQNIYNYISLTAISK 523
Qy 1170 AKTLPKGKGVVVDVAFGGAVENPEYL 1197
Db 524 LPMDSTRYQN-----SHSTAVDNPEYL 544
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:12:49 ; Search time 10.2304 Seconds
(without alignments)
5088.033 Million cell updates/sec

Title: SEQ4-5-25-14

Perfect score: 6812

Sequence: 1 MELAALCRWGLLALLPPGA.....TFKGTPTAENPEYLGIDVPV 1255

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6694	98.3	1255	1 ERB2 HUMAN	P04626 homo sapien
2	5882	86.3	1257	1 ERB2_RAT	P06494 rattus norv
3	5879.5	85.3	1254	1 ERB2_MESAU	Q06053 mesocricetu
4	3130	45.9	1210	1 EGFR HUMAN	P00533 homo sapien
5	3112	45.7	1210	1 EGFR_MOUSE	Q01279 mus musculus
6	2968.5	43.6	1308	1 ERB4 HUMAN	Q15303 homo sapien
7	2949	43.3	1308	1 ERB4_RAT	Q62956 rattus norv
8	2705.5	39.7	1167	1 XMRK_XIPWA	P13388 xiphophorus
9	2417	35.5	1342	1 ERB3 HUMAN	P21860 homo sapien
10	2350.5	34.5	1339	1 ERB3_RAT	Q62799 rattus norv
11	1931	28.3	1426	1 EGFR_DROME	P04412 drosophila
12	1749.5	25.7	634	1 ERBB_ALV	P00534 avian leuko
13	1703	25.0	604	1 ERBB_AVIER	P00535 avian eryth
14	1630	23.9	540	1 ERBB_AVIEU	P11273 avian eryth
15	1587	23.3	703	1 EGFR_CHICK	P13387 gallus gall
16	1297	19.0	1323	1 LT23_CABEL	P24348 caenorhabdi
17	1142.5	16.8	245	1 ERB2_MOUSE	P70424 mus musculus
18	724	10.6	1363	1 ILPR_BRALA	O02466 branchiosto
19	700	10.3	1382	1 INSR_HUMAN	Q06213 homo sapien
20	699.5	10.3	1607	1 MIPR_LYMST	Q25410 lymphocysta
21	695	10.2	1300	1 IRR_MOUSE	Q9T414 mus musculus
22	694.5	10.2	1372	1 INSR_MOUSE	P15208 mus musculus
23	694	10.2	1383	1 INSR_RAT	P15127 rattus norv
24	694	10.2	1477	1 HTK7_HYDAT	Q25197 hydra atten
25	687	10.1	1297	1 IRR_HUMAN	P14616 homo sapien
26	684.5	10.0	1300	1 IRR_CAVPO	P14617 cavia porce
27	645	9.5	1367	1 IGIR_HUMAN	P08069 homo sapien
28	635	9.3	1373	1 IGIR_MOUSE	Q60751 mus musculus
29	631.5	9.3	1370	1 IGIR_RAT	P24062 rattus norv
30	625.5	9.2	1390	1 INSR_AEDAE	Q93105 aedes aegypt
31	621.5	9.1	2146	1 INSR_DROME	P09208 drosophila
32	615	9.0	987	1 EPB4_HUMAN	Q4760 homo sapien
33	591.5	8.7	984	1 EPB1_CHICK	Q07494 gallus gall

RESULT 1

ID	ERB2_HUMAN	STANDARD;	PRT;	1255 AA.
AC	P04626;			
DT	13-AUG-1987 (Rel. 05, Created)			
DT	13-AUG-1987 (Rel. 05, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)			
DE	(p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell			
DE	surface receptor HER2) (MLN 19).			
GN	ERBB2 OR HER2 OR N3L OR NEU.			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=86118663; PubMed=3003577;			
RA	Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,			
RA	Saito T., Toyoshima K.;			
RT	"Similarity of protein encoded by the human c-erbB-2 gene to			
RT	epidermal growth factor receptor."			
RL	Nature 319:230-234 (1986).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=86070181; PubMed=2999974;			
RA	Coussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,			
RA	McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J.,			
RA	Francce U., Levinson A., Ullrich A.;			
RT	"Tyrosine kinase receptor with extensive homology to EGF receptor			
RT	shares chromosomal location with neu oncogene."			
RL	Science 230:1132-1139 (1985).			
RN	[3]			
RP	SEQUENCE OF 737-1031 FROM N.A.			
RX	MEDLINE=86016729; PubMed=2995967;			
RA	Semba K., Kanata N., Toyoshima K., Yamamoto T.;			
RT	"A v-erbB-related protooncogene, c-erbB-2, is distinct from the			
RT	c-erbB-1/epidermal growth factor-receptor gene and is amplified in a			
RT	human salivary gland adenocarcinoma."			
RL	Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501 (1985).			
RN	[4]			
RP	VARIANTS VAL-654 AND VAL-655.			
RX	MEDLINE=93194196; PubMed=8095488;			
RA	Ehsani A., Low J., Wallace R.B., Wu A.M.;			
RT	"Characterization of a new allele of the human ERBB2 gene by allele-			
RT	specific competition hybridization."			
RL	Genomics 15:426-429 (1993).			
CC	-!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,			
CC	ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A			
CC	POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-			
CC	ALPHA AND AMPHIREGULIN.			
CC	-!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein			
CC	tyrosine phosphate.			
CC	-!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS			
CC	(POTENTIAL).			
CC	-!- SUBCELLULAR LOCATION: Type I membrane protein.			

34	589.5	8.7	977	1	EPB4_MOUSE	Q03145 mus musculus
35	588	8.6	987	1	RET_HUMAN	P54761 mus musculus
36	588	8.6	1114	1	RET_HUMAN	P07949 homo sapien
37	584.5	8.6	976	1	EPB2_HUMAN	P29317 homo sapien
38	583.5	8.6	984	1	EPB1_RAT	P09759 rattus norv
39	579	8.5	902	1	EPBB_XENLA	Q01736 xenopus lae
40	577.5	8.5	984	1	EPB1_HUMAN	P54762 homo sapien
41	576.5	8.5	985	1	EPBA_XENLA	Q91571 xenopus lae
42	573.5	8.4	1053	1	PAK1_CHICK	Q00944 gallus gall
43	569	8.4	1068	1	PAK1_XENLA	Q91738 xenopus lae
44	567	8.3	757	1	HT16_HYDAT	P53356 hydra atten
45	563	8.3	1052	1	PAK1_MOUSE	P34152 mus musculus

ALIGNMENTS

-!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
RESIDUES (BY SIMILARITY).
-!- POLYMORPHISM: THERE ARE FOURS ALLELES DUE TO THE VARIATIONS IN
POSITIONS 654 AND 655. ALLELE B1 (654-ILE-ILE-655) HAS A FREQUENCY
OF 0.782; ALLELE B2 (654-ILE-VAL-655) HAS A FREQUENCY OF 0.206;
ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.
-!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; M11767; AAA35808.1; -
EMBL; M11761; AAA35808.1; JOINED.
DR EMBL; M11762; AAA35808.1; JOINED.
DR EMBL; M11763; AAA35808.1; JOINED.
DR EMBL; M11764; AAA35808.1; JOINED.
DR EMBL; M11765; AAA35808.1; JOINED.
DR EMBL; M11766; AAA35808.1; JOINED.
DR EMBL; M11730; AAA75493.1; -
DR EMBL; M12036; AAA35978.1; -
EMBL; X03363; CRA27060.1; -
PIR; A25491; A25491.
PIR; A24571; A24571.
DR HSSP; P11362; 1FGK.
DR Genew; HGNC:3430; ERBB2.
DR MIM; 164870; -
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; YLP motif.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Polymorphism.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 1255 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
FT DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 653 675 POTENTIAL.
FT DOMAIN 676 1255 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 720 987 PROTEIN KINASE.
FT NP_BIND 726 734 ATP (BY SIMILARITY).
FT BINDING 753 753 ATP (BY SIMILARITY).
FT ACT_SITE 845 845 BY SIMILARITY.
FT DISULFID 195 204 BY SIMILARITY.
FT DISULFID 199 212 BY SIMILARITY.
FT DISULFID 220 227 BY SIMILARITY.
FT DISULFID 224 235 BY SIMILARITY.
FT DISULFID 236 244 BY SIMILARITY.
FT DISULFID 240 252 BY SIMILARITY.
FT DISULFID 255 264 BY SIMILARITY.
FT DISULFID 268 295 BY SIMILARITY.
FT DISULFID 299 311 BY SIMILARITY.
FT DISULFID 315 331 BY SIMILARITY.
FT DISULFID 334 338 BY SIMILARITY.
FT DISULFID 511 520 BY SIMILARITY.
FT DISULFID 515 528 BY SIMILARITY.
FT DISULFID 531 540 BY SIMILARITY.
FT DISULFID 544 560 BY SIMILARITY.

FT DISULFID 563 576 BY SIMILARITY.
FT DISULFID 584 584 BY SIMILARITY.
FT DISULFID 587 596 BY SIMILARITY.
FT DISULFID 600 623 BY SIMILARITY.
FT DISULFID 626 634 BY SIMILARITY.
FT DISULFID 630 642 BY SIMILARITY.
FT MOD_RES 1139 1139 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1248 1248 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 68 68 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 124 124 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 187 187 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 259 259 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 530 530 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 571 571 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 629 629 N-LINKED (GLCNAC) (POTENTIAL).
FT VARIANT 654 654 I -> V.
FT VARIANT 655 655 /FTID=VAR_004077.
FT VARIANT 655 655 I -> V.
FT CONFLICT 1170 1170 /FTID=VAR_004078.
FT CONFLICT 1255 AA; 137909 MW; 39E9FDA04DCE962 CRC64;
SQ SEQUENCE 1255 AA; 137909 MW; 39E9FDA04DCE962 CRC64;
Query Match 98.34; Score 6694; DB 1; Length 1255;
Best Local Similarity 98.34; Pred. No. 0;
Matches 1237; Conservative 5; Mismatches 9; Indels 8; Gaps 2;
Qy 1 MELAALCRWGLLALLPPGAASSTOVCFTFNFTVSWLRVPKVSASHLE----LYQGCOV 56
Db 1 MELAALCRWGLLALLPPGAASSTOVCFTFNFTVSWLRVPKVSASHLE----LYQGCOV 56
Qy 57 QGNLELTYPNLTNLSFLQDIQEVQGVLIHQNVRQVPLQRLRIVRGTLQFEDNYALAV 116
Db 57 QGNLELTYPNLTNLSFLQDIQEVQGVLIHQNVRQVPLQRLRIVRGTLQFEDNYALAV 116
Qy 117 LDNGDPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHN 176
Db 117 LDNGDPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHN 176
Qy 177 NQALTLTIDNRSRACHPCSPCKSGSCWGESSEDCOSLTRTVCAGCACRCKGLPTDCC 236
Db 177 NQALTLTIDNRSRACHPCSPCKSGSCWGESSEDCOSLTRTVCAGCACRCKGLPTDCC 236
Qy 237 HEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPGRYTFGASCV 296
Db 237 HEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPGRYTFGASCV 296
Qy 297 TACPYNLSTDVSGCTLVCPHMQEVTAEADGTORCEKSKPCARVCYGLGMEHUREVAV 356
Db 297 TACPYNLSTDVSGCTLVCPHMQEVTAEADGTORCEKSKPCARVCYGLGMEHUREVAV 356
Qy 357 TSANIQEFACKKIFGSLAFLPESFDGDRASNTAPLOPEOLQVPEETLEETGYLYISAMP 416
Db 357 TSANIQEFACKKIFGSLAFLPESFDGDRASNTAPLOPEOLQVPEETLEETGYLYISAMP 416
Qy 417 DSLPDLISVFQNLQVIRGRILHNGAYSILTQGLGISWLGSLRSLRSLGSLALHHTHLCF 476
Db 417 DSLPDLISVFQNLQVIRGRILHNGAYSILTQGLGISWLGSLRSLRSLGSLALHHTHLCF 476
Qy 477 VHTVPMDQLFRNPQHALLHTANRPEDECVGEGLAHQLCARGHCWGPGPTQCVNCSQFLR 536
Db 477 VHTVPMDQLFRNPQHALLHTANRPEDECVGEGLAHQLCARGHCWGPGPTQCVNCSQFLR 536
Qy 537 GQECVEECRVQLGLPREYVNRHCLPCHPECOQNGSVTCFGEADOCVACAHYKDPFPC 596
Db 537 GQECVEECRVQLGLPREYVNRHCLPCHPECOQNGSVTCFGEADOCVACAHYKDPFPC 596
Qy 597 VARCPGKVPDLSYMPIWKFPDEBAGCQPCINCTHSCVDLDDKGCPCAEORASPLTSIVS 656
Db 597 VARCPGKVPDLSYMPIWKFPDEBAGCQPCINCTHSCVDLDDKGCPCAEORASPLTSIVS 656
Qy 657 AVVGILLVVLGVVFGILIKRQOKIRKYTMRLQLQETELVEPLTPSGAMPNQAMRILK 716
Db 657 AVVGILLVVLGVVFGILIKRQOKIRKYTMRLQLQETELVEPLTPSGAMPNQAMRILK 716

QY	717	ETELRKVKVLGSGAGFTYKGIWITPDGENVKIPVAIKVLRENTSPKANKETLDEAYYVAG	776
QY	717	ETELRKVKVLGSGAGFTYKGIWITPDGENVKIPVAIKVLRENTSPKANKETLDEAYYVAG	776
QY	777	VGSPVYSRLGICLTSTVQLVTQLMPPYGCLLDHHVRENRRGLSGODLLNWCQIAKGMYSYL	836
QY	777	VGSPVYSRLGICLTSTVQLVTQLMPPYGCLLDHHVRENRRGLSGODLLNWCQIAKGMYSYL	836
QY	837	EDVELVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEYHADGGKVPKMWALESIILR	896
QY	837	EDVELVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEYHADGGKVPKMWALESIILR	896
QY	897	RRFTHQSDVMSYGVTVMELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMV	956
QY	897	RRFTHQSDVMSYGVTVMELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMV	956
QY	957	KCMWIDSECPRRFRELVSERWARDQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDDMGD	1016
QY	957	KCMWIDSECPRRFRELVSERWARDQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDDMGD	1016
QY	1017	LVDAAEYLVVQQGFFCDDPAPGAGMVHRRHRSSTRSGGDLTLGLEPSEEEAPRSPLA	1076
QY	1017	LVDAAEYLVVQQGFFCDDPAPGAGMVHRRHRSSTRSGGDLTLGLEPSEEEAPRSPLA	1076
QY	1077	PSEGAGSDVFDGDLGMAAKGLQLSLTHDPSPLQRYSEDPTVPLPSTDGYVAPLTCSPQ	1136
QY	1077	PSEGAGSDVFDGDLGMAAKGLQLSLTHDPSPLQRYSEDPTVPLPSTDGYVAPLTCSPQ	1136
QY	1137	PEYUNQDVRPQPPSPREGPLPAARPAAGATILERAKTILSPKNGVYKDVAFAGGAVENPEY	1196
QY	1137	PEYUNQDVRPQPPSPREGPLPAARPAAGATILERAKTILSPKNGVYKDVAFAGGAVENPEY	1196
QY	1197	LTPGGGAAPQHPHPPAPFSPAFDNLVYWDQPPERGAPPSTFKGTPTAENPEYLGLDVVP	1255
QY	1197	LTPGGGAAPQHPHPPAPFSPAFDNLVYWDQPPERGAPPSTFKGTPTAENPEYLGLDVVP	1255

RESULT 2

ID	ERB2_RAT	STANDARD;	PRT;	1257 AA.
AC	P06494;			
DT	01-JAN-1998 (Rel. 06, Created)			
DT	15-DEC-1998 (Rel. 37, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)			
DE	(p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Epidermal growth factor receptor-related protein).			
GN	ERBB2 OR NEU.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
NCBI	TaxID=10116;			
RN	[1]			
RC	SEQUENCE FROM N.A.			
RC	TISSUE=Neuroblastoma;			
RX	MEDLINE=86118662; PubMed=3945311;			
RA	Bargmann C.I., Hung M.-C., Weinberg R.A.;			
RT	"The neu oncogene encodes an epidermal growth factor receptor-related protein.";			
RL	Nature 319:226-230(1986).			
RN	[2]			
RN	SEQUENCE OF 852-905 FROM N.A.			
RC	TISSUE=Sciatic nerve;			
RX	MEDLINE=91222560; PubMed=2025425;			
RA	Lai C., Lemke G.;			
RT	"An extended family of protein-tyrosine kinase genes differentially expressed in the vertebrate nervous system.";			
RL	Neuron 6:691-704(1991).			
RN	[3]			
RP	STRUCTURE BY NMR OF 650-668.			
RX	MEDLINE=921195181; PubMed=1346763;			
RA	Gullick W.J., Bottomley A.C., Lofts F.J., Doak D.G., Mulvey D.			


```
FT DISULFID 569 586 BY SIMILARITY.
FT DISULFID 589 598 BY SIMILARITY.
FT DISULFID 602 625 BY SIMILARITY.
FT DISULFID 628 636 BY SIMILARITY.
FT DISULFID 632 644 BY SIMILARITY.
FT MOD_RES 1141 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 188 188 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 260 260 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 631 631 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).
SQ SEQUENCE 1257 AA; 138831 MW; 6129264583011402 CRC64;

Query Match 86.3%; Score 5882; DB 1; Length 1257;
Best Local Similarity 86.3%; Pred. No. 3.6e-306;
Matches 1088; Conservative 54; Mismatches 109; Indels 10; Gaps 4;

Qy 1 MELALCRWGLLLALLPPGAASSTOVCTFNFTVSFWLRVVKVSASHLE-----LYQGCOVV 56.
Db 1 MELAAACRWGFLALLPPGAGTAGTCTGTDWK-----LRLPASPETHLDMLRHLVYQGCQVV 56
Qy 57 QGNLELTYPNALSLSFLQDIQEVQVYLIAHNOVROVPLQRLIRVGTOLFEDNYALAV 116
Db 57 QGNLELTYPNALSLSFLQDIQEVQVYLIAHNOVKKVPLQRLIRVGTOLFEDKYALAV 116
Qy 117 LDNGDPLNNTTPTV-GASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILMKDIFHK 175
Db 117 LDNRDPQDNVAASTPGRTPEGLRELQRLSLTEILKGGVLRGNPQLCYQDWLWKDVERK 176
Qy 176 NNQALTLIDNRRACHPCSPMKGRSCWGESSEDCQSILTRTVACGCCARCKGPLTDC 235
Db 177 NNQALPVDIDNRRACPPCAPACKDNHCWGESPEDCQILTGTICTSGCARCKGRPLTDC 236
Qy 236 CHEOCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTPESMPNPEGRVTFGCASC 295
Db 237 CHEOCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTPESMHNPEGRVTFGCASC 296
Qy 296 VTACPNYLSLDVSGCTLVCPPLHNOEVAEDGTQCEKCKPCARVCYGLGMEHLREVR 355
Db 297 VTTCPNYLSDEVSGCTLVCPNNQEVTAEDGTQCEKCKPCARVCYGLGMEHLRGARA 356
Qy 356 VTSANIQEFAGCKIKFSLAPLPSFDGDPASNTAPLOPEQLOVFETLEITGYLYISAW 415
Db 357 ITSDNVQDFDCKIKFISLAPLPSFDGDPSSGIAPLRPEQLOVFETLEITGYLYISAW 416
Qy 416 PDSLPDLVSFONLOVIRGRILHNGAYSILTQGLGISWGLRLSLRELGLALIHNNTHLC 475
Db 417 PDSLRDLVSFONLRIIRGRILHDGAYSILTQGLGIHSLRLSRLSGLALIHNNHLC 476
Qy 476 FVHTVPWDQLFRNPQALLHTANPEDE-CVGEGLACHOLCARGHCWGPQPTQCVCNSQF 534
Db 477 FVHTVPWDQLFRNPQALLHSGNRPEEDLCVSSGLVCNLSLCAHGCWGPQPTQCVCNSHF 536
Qy 535 LRGECEVEECRVLOGLPREYNARHCLPCHPECPQNGSVTCFPEADQCVACHYKDDP 594
Db 537 LRGECEVEECRVKWLKPREYVSDRCLPCHPECPQNSSETCFGSEADQCAACHYKDD 596
Qy 595 FCVARCPGKVPKDLVSYMPKWFPEEGACQPCPNTCHSCVDLDDKGCPCAEORASPLTSI 654
Db 597 SCVARCPGKVPKDLVSYMPKWFPEEGICQPCPNTCHSCVDLDERGCPAEORASPTFI 656
Qy 655 VSAVVGILLVVLGVWFGLIKRQKIRKYTMERLLQETELVEPLTPSGAMPNQAQMRI 714
Db 657 IATVVGILLVVLVWVVGILIKRQKIRKYTMERLLQETELVEPLTPSGAMPNQAQMRI 716
Qy 715 LKTELKRVKVLGSGAGTVYKGIWIDGENVKIPVAIKVLRNTPSPKANKEILDEAYVM 774
Db 717 LKTELKRVKVLGSGAGTVYKGIWIDGENVKIPVAIKVLRNTPSPKANKEILDEAYVM 776
Qy 775 AGVGSPPVSRLLGICLTSTVQLVTPYGLCLDHVREHRRGLSGDQLLNWCMQIAKGMS 834
```

```
Db 777 AGVGSPPVSRLLGICLTSTVQLVTPYGLCLDHVREHRRGLSGDQLLNWCMQIAKGMS 836
Qy 835 YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESI 894
Db 837 YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESI 896
Qy 895 LRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYMI 954
Db 897 LRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYMI 956
Qy 955 MVKCMWIDSECRPRFRELVSERFMRARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDM 1014
Db 957 MVKCMWIDSECRPRFRELVSERFMRARDPQRFVVIQNEIDLGPSSPMDSTFYRSLLEDDDM 1016
Qy 1015 GDVLDAEYLVPOQGFCDPPAPGAGGMVHRRSSSTRSGGDLTLGLLEPSEEEAPRSP 1074
Db 1017 GDVLDAEYLVPOQGFCDPPAPGAGGMVHRRSSSTRSGGDLTLGLLEPSEEGPRSP 1076
Qy 1075 LAPSEGAGSDVFDGLGMAAKGLQSLPTHDPSPLOQYSEDPTVLPSETDGVVAPLTCS 1134
Db 1077 LAPSEGAGSDVFDGLGMAAKGLQSLPTHDPSPLOQYSEDPTVLPSETDGVVAPLTCS 1136
Qy 1135 POPEYVQPDVRRPQPPSPREGPLPAAAPAGATLERAKTLPKGNVGVKDVAFGGAVENP 1194
Db 1137 POPEYVQSEVQPPPLTPREGPLPVPVAGATLERPKTLPKGNVGVKDVAFGGAVENP 1196
Qy 1195 EYLTPOGGAAPQHPHPPAFSPAFDNLVYMDQDPPERGAPPSTFKGTPTAENPEYLGLDVP 1254
Db 1197 EYLVPRREGTASPPHPPAFSPAFDNLVYMDQDPPERGAPPSTFKGTPTAENPEYLGLDVP 1256
Qy 1255 V 1255
Db 1257 V 1257

RESULT 3
ERR2_MESAU STANDARD; PRT; 1254 AA.
ID ERR2_MESAU
AC Q60553;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (BC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2).
DE ERBB2 OR NEU.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RT TISSUE=Nerve;
RC MEDLINE=94193007; PubMed=7908275;
RA Nakamura T., Uehijima T., Ishizaka Y., Nagao M., Arai M.,
RA Yamazaki Y., Ishikawa T.,
RT "Cloning and activation of the Syrian hamster neu proto-oncogene.";
RL Gene 140:251-255(1994).
CC -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```



```
FT DISULFID 620 628 BY SIMILARITY.
FT DISULFID 624 636 BY SIMILARITY.
FT MOD_RES 680 680 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
FT MOD_RES 1092 1092 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1110 1110 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1172 1172 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1197 1197 PHOSPHORYLATION (AUTO-, MAJOR SITE)
FT CARBOHYD 128 128 (BY SIMILARITY).
FT CARBOHYD 175 175 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 352 352 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 413 413 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 528 528 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 568 568 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 603 603 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 19 19 C -> S (IN REF. 2).
FT CONFLICT 539 539 C -> W (IN REF. 5).
FT CONFLICT 991 991 L -> F (IN REF. 4).
FT CONFLICT 1116 1117 HP -> DR (IN REF. 6).
SQ SEQUENCE 1210 AA; 134853 MW; 690E20D46DF2D2F5 CRC64;

Query Match
Best Local Similarity 45.78; Score 3112; DB 1; Length 1210;
Matches 630; Conservative 169; Mismatches 352; Indels 132; Gaps 24;

QY 11 LLLALLPPGAA-----STQVCTFNFTVFWLRVPRKVSASHLELYGCGOVQ 57
DB 14 LLLTALCAAGALBEKKVCGQTSNLTQTGFEDHFLSL-----QRMVNCVEVL 62
QY 58 GNLELTYPNLSFLQDIQVOGYVLIHAHQVQVPLQRLRIVRGTLQFEDNYALAVL 117
DB 63 GNLEITVQVNYDLSPKLTQIEVAGVYLIANTVERIPLENLQIIRGNALYENTYALAIL 122
QY 118 DNGDPLNTPVTGASPGGLRELQRLSLRILKGGVLIQRPOLYQDTILWKDI----F 173
DB 123 SN-----YGTNRTGLRELPMRNLQILIGAVRFSNPNILCNMDTITQWRDVIQNVF 172
QY 174 HKNQLALTLIDNRSBACHPCSPMKSGSRCSWESSEDCSLTRTVCAAGCA-RCKGPLP 232
DB 173 MSNWSMDL-----QSHSSCKPKDPCSPNGSCWGGEGENCQKTKIICAQOCSHRCGRSP 228
QY 233 TDCHEQCAAGCTGPKHSCDCLACILHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFG 292
DB 229 SDCHNCAAGCTGPRSDCLVCQKQFDEATCKDTCPEPLMYNPTTVQMDVNPBGKYSFG 288
QY 293 ASCVTACPNYLSLTDVSGSCTLVPLHNOEVAEDGTQRCBCKSPCARVCYGLGMEHLRE 352
DB 289 ATCKKCPRYNVVTDHSGSCVRACGPDYIEV-EEDGIRKCKKCDGPKRKCVCNGIGIEFKD 347
QY 353 VRAVTSANIOEFAGCKKIFGSLAFLPESPDGPNASNTAPLOPQLOVFTLEITGLYXI 412
DB 348 TLSINATNIKFKYCTAISGLHLVPAFGDSFTPTPDPRELELTKTKVKEITGFLLI 407
QY 413 SAWPDSLDLSVFQNLQVIRGRILHNGAYSILTLQGLISGLRLSRLRELGSGLALIHNT 472
DB 408 QAMPDNWTDLHAFENLEIRGTQKQGFQSLAVVGLNITSILGLSLKEISDGDVILSGNR 467
QY 473 HLCFVHTVPWDQFLFRNPHQALLHTANPEDECVEGEGLAGHOLCARGHCWGPGTQCVCNS 532
DB 468 NLCYANTINWKLFGTPTNQTKIMNRAEKDKAVNHVCNPLCSSESGCWPPEPRDCVSCQ 527
QY 533 QFLRGQCEVCEVQLQGLPREYNARHCLPCHPECQONGSVTCFGEADQCACAHYKD 592
DB 528 NVSRGECVBKMLLEGEPREFVENSECQCHPECUPQAMNITCTGRGPNCCQCAHYD 587
QY 593 PPFVCARCPGSKVPDLSYMPIDEPFGACQPCPCINCTHSCVDLDKDGCPAQRASPLT 652
DB 588 GPHCVKTCPAGINGENNTL-VMKYADANNVCHLCHANCTYGCAGPGQCGEVWPSGPKIP 646
QY 653 SIYSAVVGIILLVVVLGVVFGI-LIKRQKQKIRKYTWRRLLQETELVEPLTPSGAMPNQAQ 711
```

```
DB 647 STATGVGLLFTIW-VALGIGLFMRRRHIVKRTLRLLQERLEPLTPSGEAPNQA 705
QY 712 MRILKETELRKVKVLGSGAGFTYKGIWIPDGNVPIPAVKVLRENTSPKANKEILDEA 771
DB 706 LRILKETEFKIKVLGSGAGFTYKGIWIPDGNVPIPAVKELRENTSPKANKEILDEA 765
QY 772 YNAGVGSVPVSRLLGLTSTVOLVQLMPYGLCLDHYVRENRGRIGSDQLLNCMQIAK 831
DB 766 YMASVDNEFVCRLLGICLTSTVOLITQLMPYGLCLDHYVREHKDNIGSYLLNWCQIAK 825
QY 832 GNSYLEDEVLRVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEYHADGGKVPDKWMAL 891
DB 826 GMYLEDRRLVHRDLAARNVLKTPQHVKITDFGLAKLLGAEEKEYHAEGKVPDKWMAL 885
QY 892 ESILRRFRFHQSVDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDV 951
DB 886 ESILHRIYTHQSVDVMSYGVTVWELMTFGSKPYDGIIPASDISSILEKGERLPQPPICTIDV 945
QY 952 YIMVVKWMDISECRPRFRELSEFSRMDARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLE 1010
DB 946 YIMVVKWMDISDRPKFRELILEFSKMDARDPQRFVVIQ-OGDERMHLPSPTDSNFYRALMD 1005
QY 1011 DDDMGDLVDAEYLVPOQGFCEPDPAPGAGMVHHRHSSTPSGCGDLTLGLEPSEEEA 1070
DB 1006 EEDMEDVDVDAEYLVPOQGF-----NSPST----- 1031
QY 1071 PRSPLAPSGAGSDVFDGLGMAAGLQSLPDPSPQRYSEDDTVPPLPSET--DGYV 1128
DB 1032 SRTPLLSLSLATS-----NSTVACINRNGSCVKEDAFQRYSSDPTGAVTENIDDAFL 1087
QY 1129 APLTCSQPQYVNVQDVRPQPPSPREGPLPAARPAAGATLAKTLSPGKNGVVKDVFAFG 1188
DB 1088 -----PVPEYVNO-SVPKRPAGSVQNPVYHQPLHP-----APGRDLHYON--PHS 1130
QY 1189 GAVENPEYL-TQCGGAAPQHPHPPAPSPAFDNLVYWDQ-----DP-----PERG 1231
DB 1131 NAVGNPEYLNATQ-----PTCLSSGFNSPALMIQKSHQMSLDNPDYQODFFPKET 1181
QY 1232 APSTFEGTFTAENPEYLVGLDVP 1254
DB 1182 KPNGIFKG-FTAENAYLVRAPP 1203

RESULT 6
ERB4 HUMAN STANDARD; PRT; 1308 AA.
ID ERB4 HUMAN
AC Q15303;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112)
DE (p180erbB4) (Tyrosine kinase-type cell surface receptor HER4).
DE ERB4 OR HER4.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_TaxID=9606;
RP SEQUENCE FROM N.A. (ISOFORM JM-A).
RC TISSUE=Breast carcinoma;
RX MEDLINE=93189574; PubMed=8383326;
RA Plowman G.D., Culouscou J.-M., Whitney G.S., Green J.M., Carlton G.W.,
RA Foy L., Neubauer M.G., Shoyab M.;
RT "Ligand-specific activation of HER4/p180erbB4, a fourth member of the
RT epidermal growth factor receptor family.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750(1993).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM JM-A AND JM-B).
RC TISSUE=Fetal brain;
RX MEDLINE=97476287; PubMed=9334263;
RA Elenius K., Corfas G., Paul S., Choi C.J., Rio C., Plowman G.D.,
RA Klagsbrun M.;
```

"A novel juxtamembrane domain isoform of HER4/ErB4. Isoform-specific tissue distribution and differential processing in response to phorbol ester." J. Biol. Chem. 272:26761-26768 (1997).

CC -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.

CC NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN.

CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein tyrosine phosphate.

CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS (POTENTIAL).

CC -1- SUBCELLULAR LOCATION: Type I membrane protein.

CC ARE PRODUCED BY ALTERNATIVE SPLICING. THE TWO ISOFORMS DIFFER FUNCTIONALLY IN THEIR RESPONSE TO PHORBOL ESTER: JM-A IS PROCESSED BUT NOT JM-B. SO, THEY RESPECTIVELY REPRESENT CLEAVABLE AND NONCLEAVABLE FORMS OF THE RECEPTOR. BOTH ISOFORMS ARE EXPRESSED IN CEREBELLUM, BUT ONLY THE JM-B ISOFORM IS EXPRESSED IN THE HEART.

CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN BRAIN, HEART, KIDNEY, IN ADDITION TO SKELETAL MUSCLE, PARATHYROID, CEREBELLUM, PITUITARY, SPLEEN, TESTIS AND BREAST. LOWER LEVELS IN THYMUS, LUNG, SALIVARY GLAND, AND PANCREAS.

CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES.

CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; L07868; AAB59446.1; -.

DR HSP; P11362; 1FGK.

DR Genew; HGNC:3432; ERBB4.

DR MIM; 600543; -.

DR InterPro; IPR000494; EGFR_L domain.

DR InterPro; IPR000719; Euk_pkinase.

DR InterPro; IPR002174; Furin-like.

DR InterPro; IPR001245; Tyr_pkinase.

DR InterPro; IPR004019; YLP motif.

DR Pfam; PF00069; pkinase; 1.

DR Pfam; PF00757; Furin-like; 1.

DR Pfam; PF01030; Recep_L domain; 2.

DR Pfam; PF02757; YLP; 2.

DR ProDom; PD000001; Euk_pkinase; 1.

DR SMART; SMO0261; FU; 4.

DR SMART; SMO0219; TyKc; 1.

DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.

DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.

DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.

DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal; Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation; Alternative splicing

FT SIGNAL 1 25

FT CHAIN 26 1308

FT DOMAIN 26 651

FT TRANSMEM 652 675

FT DOMAIN 676 1308

FT DOMAIN 186 334

FT DOMAIN 496 633

FT DOMAIN 718 985

FT NP_BIND 724 732

FT BINDING 751 751

FT ACT_SITE 843 843

FT DISULFID 189 197

FT DISULFID 193 205

FT DISULFID 213 221

FT DISULFID 217 229

FT DISULFID 230 238

RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.

EXTRACELLULAR (POTENTIAL).

POTENTIAL.

CYTOSOLASMIC (POTENTIAL).

CYS-RICH.

CYS-RICH.

PROTEIN KINASE.

ATP (BY SIMILARITY).

ATP (BY SIMILARITY).

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

FT DISULFID 234 246 BY SIMILARITY.

FT DISULFID 249 258 BY SIMILARITY.

FT DISULFID 262 289 BY SIMILARITY.

FT DISULFID 293 304 BY SIMILARITY.

FT DISULFID 308 323 BY SIMILARITY.

FT DISULFID 326 330 BY SIMILARITY.

FT DISULFID 503 512 BY SIMILARITY.

FT DISULFID 507 520 BY SIMILARITY.

FT DISULFID 523 532 BY SIMILARITY.

FT DISULFID 536 552 BY SIMILARITY.

FT DISULFID 555 569 BY SIMILARITY.

FT DISULFID 559 577 BY SIMILARITY.

FT DISULFID 580 589 BY SIMILARITY.

FT DISULFID 593 614 BY SIMILARITY.

FT DISULFID 617 625 BY SIMILARITY.

FT DISULFID 621 633 BY SIMILARITY.

FT MOD_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).

FT MOD_RES 1188 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).

FT MOD_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).

FT MOD_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).

FT CARBOHYD 138 138 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 174 174 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 181 181 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 253 253 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 358 358 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 410 410 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 473 473 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 495 495 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 548 548 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 576 576 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 620 620 N-LINKED (GLCNAC...) (POTENTIAL).

FT VARSPLIC 626 648 NGPTSHDCIVYPTMGHSTLTPQHA -> IGSSIEDICGLMD (IN ISOFORM JM-B).

FT SQ SEQUENCE 1308 AA; 146807 MW; 5E4AE80985D88761 CRC64;

Query Match 43.6%; Score 2968.5; DB 1; Length 1308;

Best Local Similarity 45.2%; Pred. No. 6.7e-151;

Matches 612; Conservative 183; Mismatches 370; Indels 189; Gaps 30;

QY 9 WGLLLALLPPGAA-----STQVCTFNFTVSFWLRVFKVSASHLE-----LYQCGQV 56

DB 8 WVVVSLVAAGTVQPSDSQSVACGTENKLS-----SLSDLEQQYRALRKYENCEV 59

QY 57 QGNLELTYPNTASLSFLQDIQEVQGVLIHQAQVQVPLQRLRIVRGTFQDPEDNALAV 116

DB 60 MGNLEITSHEHNRDLFSRVSREVTGYVLVALNQFRYLPDENRIIRGKLYEDRYALAI 119

QY 117 LONGDPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFKHN 176

DB 120 FLNRYKDGNF-----GLQELGLKNLTEILNGVYVDQNKFLCYADTIHQDIVRNP 170

QY 177 NQALTLITDNRSRACHPCSPCKGSRGWESSEDCQSLTRTVCAAGC-ARCKGPIPTDC 235

DB 171 WPSNLTALVSTNGSGGCRCHKSCGTG-RCWGPTENHCQTLTRTVCAEQCDGRCYGPVYSDC 229

QY 236 CHEQCAAGCTGPKHSDCLACLENHSGICELHCPALVYNTDTFESMPNEGVTFCASC 295

DB 230 CHREAGGSGPKDTCDFACMNFNDSGACVTCQPTQFVYVNPPTFQLEHFNAYTYGAPC 289

QY 296 VTACPNYILSTDVGSCTLVCPHLNQEVTAEDGTCRCEKSKPCARVCYGLGMBHLREVRA 355

DB 290 VKKCPHFV-VDSSSCVRACPPSKMEV-EENGKMKCKPCTDICPKADGIGTGLSLSAQT 347

QY 356 VTSANTQEFAGCKKIFGSLAFLPESPDGPASNTAPLOPEQLOVFTLEBITGLYISAW 415

DB 348 VDSNNIDKFINCTKINGNLIFLVTHGDPYNAIEADPEKLVFRVTRVTEIFGLNIQSW 407

QY 416 PDSLPLSVFQNLQVIRGRILHNGAYSLTLQGGISWLGRLSRELGLALHNNHLC 475

DB 408 PNMWTFVSFNLTVTGGRVLYSGLSLLILKQGGITSLOFQSLKEISAGNIYTDNSNLC 467

QY 476 FVHTVPMDLFRNPHQALLHTANRPEDCVGEGEGLACHQLCARGHCGPGTQCVCNCSQFL 535

```
Db 468 YYHTINWTLTSTINQIRIVIRDRNKAENCTAEGMVCNHLCSDDGCGWGGPPDQCLSCRRFS 527
Qy 536 RGQSCVECRVQLGGLPREYNARHCLPCHPQCP-QNGSVTCFQPEADQCCACAHYKDP 594
Db 528 RGRICESNLYDGEFRFENGSIQVECDPQCKEMDGLLTCHGFGPNDCTKCSHFQGP 587
Qy 595 FCVACPSGVKPDLSYPIKFPDEBAGACQPCINCTHSCVDLDDKC-----P 643
Db 588 NCVEKCPDGLUGANSF--IFKYADPDRECHPCHPNCCTGCGNPTSHDCHYYPWTGHSTLP 645
Qy 644 AEQASPLTSISAVV--GILLVVLGVVFGILIKRROOKIRKYTWRLLOELTELVEPLTP 702
Db 646 QHAR-TPL--IAAGVIGGLFLIVVGLTFVAVYVRSIK-KKRALRFL-ETELVEPLTP 700
Qy 703 SGAMPNQAKRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPK 762
Db 701 SGTAPNQAKRILKETELRKVKVLGSGAFGVYKGIWVPEGETVKIPVAIKILNETTGP 760
Qy 763 ANKEILDEAYMAGVSGPYSTRILGICLTSTVQLVTOLMPYGCILLDHVRENRGLSGQDL 822
Db 761 ANVEFDEALIMASMDHPHLVRLGLVCLSPITQILVTOLMPHGCILLEYVHEHKONIGSQL 820
Qy 823 LNCWQIAKMSYLEDVRLVHRDLAARNVLKSNHVKITDFGLARLLDIDETEVHADGG 882
Db 821 LNCVQIAKMSYLEERLVRDLAARNVLKSNHVKITDFGLARLLDIDETEVHADGG 880
Qy 883 KVIKMALESILRRRTHOSDVMSYGVYVWELMTFGAKPYDGIPIAREIDPLEKGERLP 942
Db 881 KPIKMALECIHYRKTHOSDVMSYGVYVWELMTFGKPYDGIPTREIPDLEKGERLP 940
Qy 943 QPPTCTIDVIMVKWIMISECPRELVSERFARDPQRFVWIONED-LQAPFLD 1001
Db 941 QPPTCTIDVIMVKWIMIDADSRPKFELAAEFMRARDPQRYLVITQGGDRMKLPSPND 1000
Qy 1002 STFYRSLEDDMDGLVDAEYLVPOQGFPCDPAPGAGGMVHRHSSSTRSGGDLTL 1061
Db 1001 SKFQNLDEEDLEMDAEYLV-PQAFNIPP-----IYTSRAIDNRS-----EI 1048
Qy 1062 GLEPSEEAPRS-----PLAP-SEGAGSDVFDGDLGM 1092
Db 1049 GHSPPPAYTPMGNQFVYRDGFAAQGVSVYRAPYRPTSTIPEAPVAGATAEIFDDSCCN 1108
Qy 1093 GAAKGLQSLPDSPLORVSEDPTVPLS-----ETGYVAPLTCSPQPEYVNPQDV 1145
Db 1109 GTLRKPVAPHVQEDSSQRYASDPTVPAPERSPRGELDEGYTTPMDKPKQEYLNVE- 1167
Qy 1146 RPQPPSPREGPLPAARPAAGATLERAKTILSPKNGVVKDVAFGAVENPEYLTPQGAAP 1205
Db 1168 -----ENPFVSR-----KNGDLQ-----ALDNPEYHNASNG--- 1194
Qy 1206 QPHPPA-----FSPAFDNLVYWDQDPPERG 1231
Db 1195 ---PPKADEYVNEPLYNLTANTGLKAEYVKNNILSMPEKAKKAFDNPYWNHSLPPRS 1251
Qy 1232 A--PPSTFKGTP-----AENPEYL 1249
Db 1252 TLQHPDYLOEYSTKYFYKONGRIRPVAENPEYL 1285

RESULT 7
ERB4_RAT
ID ERB4_RAT STANDARD; PRT; 1308 AA.
AC Q62956; Q922N7;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112).
GN ERB4 OR TYRO-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
NCBI_TaxID=10116;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=98221155; PubMed=9553078;
RA Zhao Y.-Y., Sawyer D.R., Baliga R.R., Opel D.J., Han X.,
RA Marchionni M.A., Kelly R.A.;
RT "Neuregulins promote survival and growth of cardiac myocytes.
RT persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT ventricular myocytes.";
RL J. Biol. Chem. 273:10261-10269(1998).
RN [2]
RP SEQUENCE OF 848-901 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=91222560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system.";
RL Neuron 6:691-704(1991).
RN [3]
RP SEQUENCE OF 1031-1198 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Frohnert P.W., Kim S.S., Corbett J.A.;
RT "Expression of neuregulins and their putative receptors, ErbB2 and
RT ErbB3, is induced during Wallerian degeneration.";
RL J. Neurosci. 17:1642-1659(1997).
CC -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
CC 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC NTA. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
CC NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC RECEPTORS (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC HEART.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF041838; AAC08899.1; -.
CC EMBL; U52531; AAC53051.1; -.
CC HSP; P11362; IFGK.
CC InterPro; IPR000494; EGFR_L_domain.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR001245; Tyr_pkinase.
CC InterPro; IPR004019; YLP_motif.
CC Pfam; PF00757; Furin-like; 1.
CC Pfam; PF00069; pkinase; 1.
CC Pfam; PF01030; Recep_L_domain; 2.
CC Pfam; PF02757; YLP; 2.
CC PRINTS; PR00109; TYRKINASE.
CC PRODOM; PD000001; Euk_pkinase; 1.
CC SMART; SM00261; FU; 4.
CC SMART; SM00219; TyrKC; 1.
CC PROSITE; PS00107; PROTEIN KINASE ATP; 1.
CC PROSITE; PS50011; PROTEIN KINASE_DOM; 1.
CC PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
RN SIGNAL 1 25 POTENTIAL.
```


FT CHAIN 26 1308 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
FT DOMAIN 26 651 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 652 675 POTENTIAL.
FT DOMAIN 676 1308 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 186 334 CYS-RICH.
FT DOMAIN 496 633 PROTEIN KINASE.
FT DOMAIN 718 985 ATP (BY SIMILARITY).
FT NP BIND 724 732 ATP (BY SIMILARITY).
FT BINDING 751 751 ATP (BY SIMILARITY).
FT ACT_SITE 843 843 BY SIMILARITY.
FT DISULFID 189 197 BY SIMILARITY.
FT DISULFID 193 205 BY SIMILARITY.
FT DISULFID 213 221 BY SIMILARITY.
FT DISULFID 217 229 BY SIMILARITY.
FT DISULFID 230 238 BY SIMILARITY.
FT DISULFID 234 246 BY SIMILARITY.
FT DISULFID 249 258 BY SIMILARITY.
FT DISULFID 262 289 BY SIMILARITY.
FT DISULFID 293 304 BY SIMILARITY.
FT DISULFID 308 323 BY SIMILARITY.
FT DISULFID 326 330 BY SIMILARITY.
FT DISULFID 503 512 BY SIMILARITY.
FT DISULFID 507 520 BY SIMILARITY.
FT DISULFID 523 532 BY SIMILARITY.
FT DISULFID 536 552 BY SIMILARITY.
FT DISULFID 555 569 BY SIMILARITY.
FT DISULFID 559 577 BY SIMILARITY.
FT DISULFID 580 589 BY SIMILARITY.
FT DISULFID 593 614 BY SIMILARITY.
FT DISULFID 617 625 BY SIMILARITY.
FT DISULFID 621 633 BY SIMILARITY.
FT MOD_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1188 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 138 138 N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD 174 174 N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD 253 253 N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD 358 358 N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD 410 410 N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD 473 473 N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD 495 495 N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD 548 548 N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD 620 620 N-LINKED (GLCNAC.) (POTENTIAL).
FT CONFLICT 1062 1062 S -> N (IN REF. 3).
FT CONFLICT 1080 1082 PYT -> SYR (IN REF. 3).
SQ SEQUENCE 1308 AA; 146957 MW; D944BB096A08B41 CRC64;

Query Match 43.3%; Score 2949; DB 1; Length 1308;
Best Local Similarity 45.0%; Pred. No. 7.3e-150;
Matches 610; Conservative 190; Mismatches 379; Indels 176; Gaps 30;

Qy 1 MELA-ALCRWGLLL--ALLPPGAASQVCTFNFTSVFWLRVPKVSASHLE-----L 49
Db 1 MKLATGLVWGSLLVAARTVQPSASQSVCACTENKLS-----SLSDLEQQYRALRKY 52
Qy 50 YQGCVQVGNLELYLPNLSLFLQDIQVQGVLYIAHNOVROVPLORLIRVGTOLFE 109
Db 53 YENCEVWGNLEITSIEHNRLSLSIREVTGVLYVALNQFRLPLENRIIRGTLYE 112
Qy 110 DNVALAVLDNGDPLNNTPVTGASPGGLRELQRLSLTEILKGGVLIORNPOLCYQDTILW 169
Db 113 DRYALALFLNRKGNF-----GLQELGLKNLLEILNGGVYVQNKFLCYADTIHW 163
Qy 170 KDIFHKNQLALTIDNRSRACHPCSPMKGSRGWGESSDCQSLTRTVCAGGC-ARCK 228
Db 164 QDIVRNWPNSNWLTVSTIGSGGRCCHKSCGTG-RCWGPTEHNCQTLRTVCAGCDGRCY 222
Qy 229 GPLPTDCHEOCAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESHPNPEGR 288
Db 223 GPVYSDCCRCAGCGSGPKDTCDFACMNFNDSGACVTCQCPQTEFVYNPNTTTFQLEHFNNAK 282

Qy 289 YTEGASCVTACPNYLLSTDVSGCTCLPLHNOQVTAEDGTQRCCKSPKCARVCYGLGME 348
Db 283 YTYGACVVKCPHFV-VDSSCVCRACPPSSKMEV-EENGIKMCKPCTDICKACDGIITG 340
Qy 349 HLEVRATVSANTQEFAGCKKIFGSLAFLPESPDGPASNTAPLOPELOVFTLEBITG 408
Db 341 SLMSAQTVDSNIDKFINCTKINGNLIFLVGTIGHDPYNAIDAIDPEKLNVTFTVRBITG 400
Qy 409 YLYISAWPDSLPLSVFQNLQVIRGILHNGAVSLTQGLGISWGLRSILRELSGLALI 468
Db 401 FLNIQTWPPNMTDFSVFSNLVTIGGRVLYSGLLLIKQOGITSLQFSKESAGNIYI 460
Qy 469 HHNTHLCFVHTVPWDQLFRNPHQALLHTANRPEDECVGEGSLACHQLCARGHCPGPTQC 528
Db 461 TDNSNLCCYHTINWTTLSFVINOIIVIRONRAENCTAEGMVCNHLCSNDGCGWPGPDQC 520
Qy 529 VNCQFLRGQECVECRVLOGLPREVYNARHCLPCHPEQOP-QNGSVTCFPGPADQCVCAC 587
Db 521 LSCRRFSRGKICIESCNLYDGBEFEPENGSIQCEVDSQCEKMEDEGLLTCHGPGDNCCTK 580
Qy 588 AHYKDPFPFCVARCPSGVKPDLSYMPIWKEPDEBEGACQPCPINCTHSCVDLDDKGC----- 642
Db 581 SHFKGQPCNCEKCPDLQGANSE--IFKYADQDRECHPCHPNCTQGCGTSHDCIYYPW 638
Qy 643 -----PAEQRASPLTSIVSAVV-GILLVVVLGVWFGLIKIRKQKIRKYMRLLOETE 695
Db 639 TGHSTLPOHAR-TPL--IAAGVIGGLFILVIMALTAVYVRRKSIIK-KKALARRFL-ETE 693
Qy 696 LVEPLTPSGAMPNQAQMRILKETELRKVKVLGSGAGFTVYVKGWIIDGENVKIIPVAIKVL 755
Db 694 LVEPLTPSGTAPNQAQRLILKETELRKVKVLGSGAGFTVYVKGWPEGETVKIIPVAIKIL 753
Qy 756 RENTSKANKEILLDEAYVMAGVCSPIVSRLLGICLTSTVOLVTLQMPYGCCLLDHVRNKG 815
Db 754 NETTGPKANVFDEALIMASVDHPLHVLRLGVLCLPTQLVTLQMPHGCCLLYVHHKD 813
Qy 816 RLGSQDLLNWCMIQAKGMSYLEDLVLRDLAARNVLKSPNHNKVTDFGLARLLDIDET 875
Db 814 NIGSOLLNWCVOIAGKMWLEERLVRDLAARNVLKSPNHNKVTDFGLARLLLEGDEK 873
Qy 876 EYHADGKVPKIMWALESILRRRPTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLL 935
Db 874 EYNADGKMKIKWMALECIHYRKFTHQSDVMSYGVTVWELMTFGKPKPYDGIPTREIPDLL 933
Qy 936 EKERLPPQPICTIDVYMLMVKCMWIDSECRFRELVSFSSRMARDPOSFVIVIONED-L 994
Db 934 EKERLPPQPICTIDVYMLMVKCMWIDADSRPKFELAAEFSSRMARDPOQLVLIQGDRLM 993
Qy 995 GPASPLDSTFYRSLLEDMDGLVDAAEYLVPOQGFPCDP----- 1035
Db 994 KLFPNDSEKFFONLLDEEDLEMDMAEYLVLP-QAFNIPPIYTSRTRIDSNRSEIGHSP 1052
Qy 1036 ----APGAGMWHHRSSSTRSGGDLTLGLBPSBEEAPRSPPLASBEGAGSDVFDGDLG 1091
Db 1053 PPATYPMGSGQFVYQGGFATQGG---MPMPYATTSTIPEAPVA--QGATAEMFDDSCC 1107
Qy 1092 MGAAGLQSLPHTDPSLQRYSEDPTVPLPS-----ETDGYVAPLTCSPQBYVYNQPD 1144
Db 1108 NGTLRKPVPVPHVQEDSSTORYSADPTVFAPERNPRAELDEEGYMTMHDKPKQBYLNPVE 1167
Qy 1145 VRQPPSPRGPLPAARPAAGATLERAKTLPSPGNKGVKDVFAFGGAVENPEYLTPQGGAA 1204
Db 1168 -----ENPFVSR-----KNGDLQ-----ALDNEYHSASSG-- 1194
Qy 1205 PQHPPPA-----FSPAFDNLYYWDQDPPER 1230
Db 1195 ----PPKADEYVNEPLYLNTFTNALGNAEYMKNSLLSVPEKAKAFDNPYWNHSLPPR 1250
Qy 1231 GA--PSTFKGTPT-----AENPEYL 1249
Db 1251 STLQHPDYLOEYSTKYFYKONGRIRPIVAENPEYL 1285

RESULT 8
XMRK_XIPWA
ID XMRK_XIPWA STANDARD; PRT; 1167 AA.
AC P13388;
DT 01-JAN-1990 (Rel. 13, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Melanoma receptor protein-tyrosine kinase precursor (EC 2.7.1.112).
GN XMRK OR TU.
OS Xiphophorus maculatus (Southern platyfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthopterygia; Acanthopterygii; Percomorpha; Atherinomorpha;
OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
OX NCBI_TaxID=8083;
[1]
SEQUENCE FROM N.A.
RX MEDLINE=90015140; PubMed=2797166;
RA Wittbrodt J., Adam D., Malitschek B., Maueier W., Raulf F.,
RA Telling A., Robertson S.M., Scharltl M.;
RT "Novel putative receptor tyrosine kinase encoded by the melanoma-
RT inducing Tu locus in Xiphophorus.";
RL Nature 341:415-421(1989).
[2]
RN REVISION TO 515.
RA Scharltl M.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
EMBL; X16891; CAA34770.2; -;
PIR; S06142; S06142.
HSP; P11362; IFGK.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR PRINTS; PR001030; Recep_L domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 5.
DR SMART; SM00220; S_TKC; 1.
DR SMART; SM00219; TYRK; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
DR PROSITE; PS00011; PROTEIN KINASE_DOM; 1.
DR Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.
FT SIGNAL 1 25
FT CHAIN 26 1167 MELANOMA RECEPTOR PROTEIN-TYROSINE
FT KINASE.
FT DOMAIN 26 642 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 643 665 POTENTIAL.
FT DOMAIN 666 1167 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 710 977 PROTEIN KINASE.
FT NP_BIND 716 724 ATP (BY SIMILARITY).
FT BINDING 743 743 ATP (BY SIMILARITY).
FT ACT_SITE 835 835 BY SIMILARITY.

FT DISULFID 195 204 BY SIMILARITY.
FT DISULFID 199 212 BY SIMILARITY.
FT DISULFID 220 228 BY SIMILARITY.
FT DISULFID 224 236 BY SIMILARITY.
FT DISULFID 237 245 BY SIMILARITY.
FT DISULFID 241 253 BY SIMILARITY.
FT DISULFID 256 265 BY SIMILARITY.
FT DISULFID 269 296 BY SIMILARITY.
FT DISULFID 300 311 BY SIMILARITY.
FT DISULFID 315 330 BY SIMILARITY.
FT DISULFID 333 337 BY SIMILARITY.
FT DISULFID 504 513 BY SIMILARITY.
FT DISULFID 508 521 BY SIMILARITY.
FT DISULFID 524 533 BY SIMILARITY.
FT DISULFID 537 553 BY SIMILARITY.
FT DISULFID 556 569 BY SIMILARITY.
FT DISULFID 560 577 BY SIMILARITY.
FT DISULFID 593 615 BY SIMILARITY.
FT DISULFID 618 626 BY SIMILARITY.
FT DISULFID 622 634 BY SIMILARITY.
FT CARBOHYD 114 114 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 144 144 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 201 201 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 356 356 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 365 365 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 398 398 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 501 501 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 621 621 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1167 AA; 129934 MW; 4793E0749DC1D55A CRC64;
Query Match 39.7%; Score 2705.5; DB 1; Length 1167;
Best Local Similarity 46.2%; Pred. No. 6 1e-137;
Matches 563; Conservative 156; Mismatches 372; Indels 128; Gaps 24;
QY 48 ELYQGVQVQGNLEIYLPNTNASLSPLODIEVQGVYLAHQVROVPLQRLIRIVRGTL 107
DB 51 KMYSGCNVYLENIYTYQENQDLSFLQSIQEVGGVLIAMNEVSTIPLVNLRLIRGQNL 110
QY 108 FEDNYALAVLDGDPNNTPTVTGASPGRLRELQRLSLTEILKGGVLIQORNPOLCYQDTI 167
DB 111 YEGNFTLLVMSYQK-NPSSP--DVYQVGLKQLSLNTEILSGGVKVSHPNLLCNVETI 167
QY 168 LWKDIFHKNNQLALTLIDTNRGRACHPCSPMCKGRGCGESSEDCQSLTRTVTCAGC-AR 226
DB 168 NWDIVDKTSNPTMNLIPAFERQCKQCDHGVCNSCWAPGPHCKQFKLILCAEQCNR 227
QY 227 CKGPLPTDCCHQCAAGCTGPKHSDCLACLFHNSHGICELHCPALVYNTDTFESMPNPE 286
DB 228 CRGPKPIDCCNEHCAGGCTGPRATDCLACRDFNDGDKTCTPPPKIYDIVSHQVVDNPN 287
QY 287 GRYTFGASCVTACPNYLSLTDVGSCTLVCLHQNQEVTAEDGTQRCCKSKPCARVCYGLG 346
DB 288 IKYTFGACVCPCPSNYVYTE-GACVRSACAGMLEVD-ENGKRSCKPCDGVCPKVCDDGIG 345
QY 347 MEHLREVRAVTSANTQEFAGCKKIFGSLAFLPESDGDPSANTAPLQPOLQVFELEBI 406
DB 346 IGSLSNTAVNSTNRSFNCTKINGDIILNRSFEGDPHYKIGTMDPHELNWLTIVTKBI 405
QY 407 TCYLYISAMPDLSPLSVFQNLQVIRGIRLHNGAYS-LTLQGLIGISWLGLSLRLGSLG 465
DB 406 TCYLVIMWMPENMTSLVFQNLLEIRGTRTFSGRGSFVVVQVVRHLQWLGLSLRSLKEVSACN 465
QY 466 ALIHNTHLCHFVHTVPWDLFRNPQALLHTANRDEDECVCEGLACHQLCARGHCWGP 525
DB 466 VILKNTLQRYANTINWRRLERSEDSQSEYDART-----ENQCTNNECESDGCWGP 518
QY 526 TCVCNCSQFLRQCEVCECRVLQGPREVYNARHCLPCHPECPQNGSVTCGPEADQCV 585
DB 519 TMCVSLHVDGRGRCVASCNLLQGBPREAQVDGRVCQCHQCLVDTSLTYCGPGPANC 578
QY 586 ACAHYKDPFFVARCPSPGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKCPAE 645

```
Db 579 KSAHFQDPQPCIPCPHGILGDGDTL-IMKYADKMGQCQCHQNCQTCGSCGSLGCRGD 637
Qy 646 QRASPLTSIVSAGVGLLVVLGVVFGILIKRQOKKRYKTMRLLOETBELVPLTPSGA 705
Db 638 -IVSHSLAVGLVSLGTLITVIVALLIVLLRRRIK-RKRTIRCLLQEKELVPLTPSGQ 695
Qy 706 MPNOAQRILKETELRVKVLGSAFGTVYKGIWIPDGENVKIPVALKVLRENTSPKANK 765
Db 696 APNOAFRIILKETEFKDKRVLGSAFGTVYKGLWNPDPGENIRIPVALKVLRENTSPKQ 755
Qy 766 EILDEAYVMAVGSPPYVSRLLIGICTSTVQLVLTQLMPEYGLLDHVRNENRGLASQDLN 825
Db 756 EVLDEAYVMAVDHPVPHVCRLLIGICTSAVLQVLTQLMPEYGLLDVYRQHERICQW 815
Qy 826 CMQIAKMSYLEDLVRLVHRLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVP 885
Db 816 CVQIAKMSYLEERHLVHRLAARNVLLKNPNHVKITDFGLSKLLTADKEXYQADGKVP 875
Qy 886 IKWMALESILRRRTHQSDVMSYGVTVWELMTGCAKPYDGIIPAREIPDLLEKGERLP 945
Db 876 IKWMALESILQWTHQSDVMSYGVTVWELMTFGSKPYDGIIPAKETIASVLENGERLP 935
Qy 946 ICTIDVYMWKMWIDSECRPRELVSFMRMARDPQRFVWQNEEDLGPASPLDSTFY 1005
Db 936 ICTIEVYMWILKMWIDPSSPRELVSFMRMARDPQRFVWQNEEDLGPASPLDSTFY 992
Qy 1006 RSLLEDDMDGLVDAEYLVVQOQFCFPAPGAGVHHRHSSSTRSGGDLTGLGLEP 1065
Db 993 RSLLSDD--DVEDADEYLLPYKRI-----NRQGS----- 1020
Qy 1066 SEEAAPSPLAPSGAGSDVFDGLGMAKGLQSLTHDPSPLORYSEDPV-PLPSET. 1124
Db 1021 -----EPCIPPTGH-----PVRENSITLNRISDPTQNALEKDL 1053
Qy 1125 DGYVAPLTCSPQPVYVQVVRPQ-----PSPRE-----GPLP-AARPAGATLBR 1171
Db 1054 DGH-----EYVNPQSETSSRLSDIYNPNVEDLTDGWPVSLSSQEAETNFSPE 1103
Qy 1172 TLPSPKNGVNDVFAFGAVENPYLTPQGAAPQHPHPAFNLYWDDPPERG 1231
Db 1104 YLNTNQNSL--PLVSSGSMDDPDY---QAG-----YQAAF-----LPQTG 1138
Qy 1232 APPSTFKGTPTAENPEYLG 1250
Db 1139 ALTNGMFLPAENLEYLG 1157

RESULT 9
ERB3 HUMAN
ID -ERB3 HUMAN STANDARD; PRT; 1342 AA.
AC P21860;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (c-erbB3) (tyrosine kinase-type cell surface receptor HER3).
GN ERB3 OR HER3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90083234; PubMed=2687875;
RA Kraus M.H., Issing W., Miki T., Popescu N.C., Aaronson S.A.;
RT "Isolation and characterization of ERB3, a third member of the
RT ERB3/epidermal growth factor receptor family: evidence for
RT overexpression in a subset of human mammary tumors."
RL Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90311312; PubMed=2164210;
```

```
RA Plowman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
RA Todaro G.J., Shoyab M.;
RT "Molecular cloning and expression of an additional epidermal growth
RT factor receptor-related gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
RN [3]
RP SEQUENCE FROM N.A. (SHORT FORM).
RC TISSUE=Placenta;
RX MEDLINE=93282822; PubMed=7685162;
RA Katoh M., Yazaki Y., Sugimura T., Terada M.;
RT "c-erbB3 gene encodes secreted as well as transmembrane receptor
RT tyrosine kinase.";
RL Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTK.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC SECRETED (SHORT FORM).
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
CC SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC -1- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M29366; AAA35790.1; -
DR EMBL; M34309; AAA35979.1; -
DR EMBL; S61953; AAB26935.1; -
DR PIR; A36223; A36223.
DR HSP; P11362; LFQK.
DR Genew; HGNC:3431; ERBB3.
DR MIM; 190151; -
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; kinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; FALSE_NEG.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Alternative splicing.
FT SIGNAL 1
FT CHAIN 20 1342
FT DOMAIN 20 643
FT TRANSMEM 644 664
FT DOMAIN 665 1342
FT DOMAIN 709 966
FT NP_BIND 715 723
FT BINDING 742 742
FT ACT_SITE 834 834
FT DISULFID 186 194
FT DISULFID 190 202
FT POTENTIAL.
FT RECEPTOR PROTEIN-TYROSINE KINASE ERBB-3.
FT EXTRACELLULAR (POTENTIAL).
FT POTENTIAL.
FT CYTOPLASMIC (POTENTIAL).
FT PROTEIN KINASE.
FT ATP (BY SIMILARITY).
FT ATP (BY SIMILARITY).
FT BY SIMILARITY.
FT BY SIMILARITY.
```


RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=96096535; PubMed=8522190;
RA Hellyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Koland J.G.;
RT "Cloning of the rat ErbB3 CDNA and characterization of the
RL recombinant protein.";
RN Gene 165:279-284 (1995).
[2]
RP REVISIONS TO 85; 513 AND 565.
RA Hellyer N.J., Koland J.G.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE OF 922-1097 FROM N.A.
RP STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Frohnert P.W., Kim S.S., Corbett J.A.;
RT "Expression of neuroligins and their putative receptors, ErbB2 and
RL ErbB3, is induced during Wallerian degeneration.";
RN J. Neurosci. 17:1642-1659 (1997).
CC -!- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTK.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U29339; AAC28498.2; -;
CC EMBL; U52530; AAC3050.1; -;
CC HSP; P11362; IFCG.
CC InterPro; IPR000494; EGFR_L domain.
CC InterPro; IPR000719; Euk_Pkinase.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR001245; Tyr_pkinase.
CC Pfam; PF00069; pkinase; 1.
CC Pfam; PF01030; Recep_L domain; 2.
CC PRINTS; PR00109; TYRKINASE.
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00261; FU; 5.
CC SMART; SM00219; TyrKc; 1.
CC PROSITE; PS00107; PROTEIN KINASE ATP; FALSE_NEG.
CC PROSITE; PS00109; PROTEIN KINASE TYR; 1.
CC PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
FT Transferrase; tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 19
FT CHAIN 20 1339
FT DOMAIN 20 643
FT TRANSMEM 644 662
FT DOMAIN 663 1339
FT DOMAIN 183 259
FT DOMAIN 707 964
FT NP_BIND 713 721
FT BINDING 740 740
FT ACT_SITE 832 832
FT DISULFID 186 194
FT DISULFID 190 202
FT DISULFID 210 218
FT DISULFID 214 226
FT DISULFID 227 235

FT DISULFID 231 243 BY SIMILARITY.
FT DISULFID 246 255 BY SIMILARITY.
FT DISULFID 259 286 BY SIMILARITY.
FT DISULFID 290 301 BY SIMILARITY.
FT DISULFID 305 320 BY SIMILARITY.
FT DISULFID 323 327 BY SIMILARITY.
FT DISULFID 500 509 BY SIMILARITY.
FT DISULFID 504 517 BY SIMILARITY.
FT DISULFID 520 529 BY SIMILARITY.
FT DISULFID 533 549 BY SIMILARITY.
FT DISULFID 556 573 BY SIMILARITY.
FT DISULFID 576 585 BY SIMILARITY.
FT DISULFID 589 610 BY SIMILARITY.
FT DISULFID 613 621 BY SIMILARITY.
FT DISULFID 617 629 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 408 408 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 437 437 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 469 469 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 522 522 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 566 566 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 616 616 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 1028 1028 L > P (IN REF. 3).
SQ SEQUENCE 1339 AA; 147545 MW; 0AA5F2402BFDFF1E CRC64;

Query Match 34.5%; Score 2350.5; DB 1; Length 1339;
Best Local Similarity 40.8%; Pred. No. 5.9e-118;
Matches 522; Conservative 170; Mismatches 419; Indels 167; Gaps 34;

QY 5 ALCRWGLLLALLPPGAASSTQCTFNFTSFVLRVPSVSHLELYOGCVVQGNLELY 64
DB 27 AVCP-GTLNGLSVTGDADNQYTL-----YKLYEKCEVVMGNLEIVL 67
QY 65 LPTNASLSFDIQIOEVQGYVLIHAHNOVRQVPLQRLRIVRGTLQFDNYALAVLDNGDPLN 124
DB 68 TGNADLSFLQWREVTGTVLVAMNFSVLPNLRVRCVQYDGFVFM-----LN 122
QY 125 NTPVTGASPGGLREIQLRSLTEILKGVLIQRNPQLCYODTILMKDIFHNQNLATLI 184
DB 123 YNT-----NSSHALRQLKFTQLTEILSGVYIEKNDKLCMDTIDWRDIVR---GAEIV 175
QY 185 DTNRSRACHPCSPMKSGSCWGESSEDCOSLTITVCAGGC-ARCKGLPTDCCHCEQCAAG 243
DB 176 VRKNGANCPCPCHEVCVGK-RCWGPGDDQLTKTTCAPQCNGRCFGPNPQCCHDECAGG 234
QY 244 CTGPKGSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPNY 303
DB 235 CSGPQDTDCFACRRFNDSGACVPCRCPELVYKLTITQLEPNPHTKYQYGGVGVASCPHNF 294
QY 304 LSTDVGSCTLVCPHMQEVTAEQGTQCEKSKPCARVCYGLGMEHLREVRVTSANIQE 363
DB 295 V-VQQTFCVRACPPDKMEVD-KHGLKMCPCGGGLCPKACEGTGSG--SRYTQTVDSNIDG 350
QY 364 FACKKIFGSLAFIPESFDGDPASNTAPLOEQLOVPEETLEETGLYLYISAWPDSLPLS 423
DB 351 FVNCTKLGLNLDLITGLNVDYDHPKIPALDPEKLNVFRTVREITGYLNIOSWPPHMFNS 410
QY 424 VFQNLQVIRGRILHNGAYS-LTLQGLGISWGLRSRLRELGSGLALIHNTLHLCFVHTVPV 482
DB 411 VFSNLITIGRSLYNGRFSLLIMKLNVTSLGFRSLKEISAGRVYISANQOLCYHHSNLW 470
QY 483 DQLFNPHQALLHTA-NRPDECVGEGLACHQICARGHCWGPGPTQCNCQSQFLRGECV 541
DB 471 TLLRGPSERLDIKYDRPLGCECLAEGKVCDDLPCSSGGCWGPGPGQCLSRNYSREGVCV 530
QY 542 EECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEPEADQCACAHYKDPPECVACRP 601
DB 531 THCNFLQGEPRFVHEAQCFSCHEPCLPMEGTSTCNSSGSDACARCAHDFGPHCVNSCP 590
QY 602 SGVKPDLSPYMPIWKFPEDEGACQPCPINCTHSC--VDLDDKGPAPQASPLTSIVSAV 659

Db 591 HGILG - AKGIYKPAQNECRCHENCQTGCGNPELQDCGAEVLMKSPHLVIAVT 648
Qy 660 GILVAVVGVVFGILIKRQOKIR-KYTMRLLOETELVEPLTPSGAMPNQAQMRILKET 718
Db 649 G-LAVILMILGSGFLYWRGRIRIONKRAMRYLERSIESIPLDP-S-EKANKVLARIFKET 705
Qy 719 ELRKVKVLGSGAFGVYKGIWIPGENVKIPVAIKVIRENTSPKANKEILDEAYVMAGVG 778
Db 706 ELRLKVLGSGVFGVTHKGIWIPGESIKIPVCIKVIEDKSGRQSFQAVTHMLAVGSLD 765
Qy 779 SPVYSLGLGLCLSTVOLVQLMPEYGLDHWRENRGRGLSGODLLNMCQIAKMSVLED 838
Db 766 HAHILVRLGLCPGSSLSQVLYQLPLGSLLDHVKQHRETLGQLLLNMGVQIAKMYILEE 825
Qy 839 VRLVHRLAARNVLAKSPNPKIKIDFGLARLLDIDETEHYADGKGVPIKMWALESIILRRR 898
Db 826 HSMVHRDLALRNVLKSPSQVQVADFGVADLLPDDKQLLHSEAKTIKMWALESIHFGK 885
Qy 899 FTHQSDVMSYGVTVWELMTGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVVKC 958
Db 886 YTHQSDVMSYGVTVWELMTGAEPYAGLRLAEIPDLLEKGERLAQPOICTIDVYIMVVKC 945
Qy 959 WMIDSECRPFRELVSFMSWARDPQRFVIONEDLGPASPLDSTFFRSILLEDGMDGLV 1018
Db 946 WMIDENLRPTFKELANEFTRWARDPPRYLVIKRAS-GFGTGP--PAAPSPVLTTKEL--- 998
Qy 1019 DAEYLVPOQGFCCPDAPGAGGMVHRHRSSTRSRGGDLTLGLEPSEE----- 1068
Db 999 -QEALEPEL-----DLDLDLEEEGLATSLGSL 1028
Qy 1069 -----EAPRSLAPSEG-----AGSDVFDGLGMGAQGLQSLTHDPSPQLQ 1110
Db 1029 SLPTGTLTRPGQSLLSPSSGYPMNMOSLIGEACLDSAVLGREQFSPRLSLH-PIPRG 1087
Qy 1111 RYSEDPVPLPSETDGVY----APL-----TC-----SPOPE-----YVNPQDVRPQP 1150
Db 1088 R-----PASESEGHVTSGEAELOBKVSVCRSRSRSPRPRGDSAYHSQRHSLTPV 1140
Qy 1151 SPREGP-----LPAARPAAGATLERAKTLSP-GKGVW-----KDVFAFGGAVE 1192
Db 1141 TPLSPGLEBEDGNGYVMPDTHLRGASSREGTILSSVGLSVLTGTEEDD-----E 1192
Qy 1193 NPEYLTPOGGAAPQHPPP 1210
Db 1193 EYEMNRKRGSP-PRPP 1209

RESULT 11

EGFR DROME
ID EGFR DROME STANDARD; PRT; 1426 AA.
AC P0412; O61601; Q9W2G0; P81868;
DT 13-AUG-1987 (Rel. 05, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Egfr)
DE (Gurken receptor) (Torpado protein) (Drosophila relative of ERBB).
GN EGFR OR TOP OR C-ERBB OR DER OR CG10079.
OS Drosophila melanogaster (Fruit fly)
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC -Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
[1]
RN
RP
PP
RX MEDLINE=94350209; PubMed=8070664;
RA Clifford R., Schupbach T.;
RT "Molecular analysis of the Drosophila EGF receptor homolog reveals that several genetically defined classes of alleles cluster in subdomains of the receptor protein."
RL Genetics 137:531-550(1994).
RN
RP REVISIONS.

RA Clifford R., Schupbach T.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
[3]
RP
RX
RX MEDLINE=85124611; PubMed=2982499;
RA Livneh E., Glazer L., Segal D., Schlessinger J., Shilo B.-Z.;
RT "The Drosophila EGF receptor gene homolog: conservation of both hormone binding and kinase domains."
RL Cell 40:599-607(1985).
[4]
RP
RX
RX STRAIN=Oregon-R; TISSUE=Embryo;
RX MEDLINE=87002474; PubMed=3093080;
RA Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;
RT "Alternative 5' exons and tissue-specific expression of the Drosophila EGF receptor homolog transcripts."
RL Cell 46:1091-1101(1986).
[5]
RP
RX
RX SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION ANALYSIS.
RX MEDLINE=99102120; PubMed=9882502;
RA Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;
RT "Several levels of EGF receptor signaling during photoreceptor specification in wild-type, Ellipse, and null mutant Drosophila."
RL Dev. Biol. 205:129-144(1999).
[6]
RP
RX
RX SEQUENCE FROM N.A. (ISOFORM TYPE I).
RX STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D., Ananthan P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F., George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N., Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X., Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D., Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G., Adair J.F., Agbayani A., An H.-J., Andrews-Frannkoch C., Baldwin D., Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., Beeson K.Y., Benton M.P., Bereman B.P., Bhandari D., Bolshakov S., Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P., Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I., Cherry J.M., Cawley S., Dahike C., Davenport L.B., Davies P., de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W., Fogle C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K., Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C., Jaiswal M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X., Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A., Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L., Palazzolo R., Pittman K.A., Nixon K., Nusskern D.R., Pacleb J.M., Palazola D.M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G., Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H., Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T., Spier E., Spradling A.C., Stappleton M., Strong R., Sun E., Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X., Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J., Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
[7]
RP
RX
RX SEQUENCE OF 959-1073 FROM N.A.
RX STRAIN=Daekwanyeong;
RX MEDLINE=85137938; PubMed=2983232;
RA Wadsworth S.C., Vincent W.S. III, Bilodeau-Wentworth D.;
RT "A Drosophila genomic sequence with homology to human epidermal

RT growth factor receptor.";
 RL Nature 314:178-180(1985).
 RN [8]
 RP SEQUENCE OF 1133-1137; 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION
 ANALYSIS.
 RX MEDLINE=92038942; PubMed=936959;
 RA Raz E., Schejter E.D., Shilo B.Z.;
 RT "interallelic complementation among DER/flb alleles: implications for
 the mechanism of signal transduction by receptor-tyrosine kinases.";
 RL Genetics 129:191-201(1991).
 RN [9]
 RN REVIEW.
 RP MEDLINE=97248481; PubMed=9094709;
 RA Perrimon N., Perkins L.A.;
 RT "there must be 50 ways to rule the signal: the case of the Drosophila
 EGF receptor.";
 RL Cell 89:13-16(1997).
 CC -!- FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARGOS,
 WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF-
 MAPK PATHWAY. INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS.
 CC CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE
 DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL
 POLARITIES OF THE OOCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE
 ESTABLISHMENT OF VENTRAL CELL FATES, MAINTENANCE OF AMNIOSESOSA
 AND VENTRAL NEUROECTODERMAL CELLS, GERM BAND RETRACTION, CELL FATE
 SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF
 CUTICLE.
 CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 tyrosine phosphate.
 CC -!- SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I
 MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE
 PROTEIN.
 CC -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS, TYPE I (SHOWN HERE), TYPE II AND
 TYPE III; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -!- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED IN EMBRYOS. IN LARVAE,
 UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANLAGEN OF
 TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST
 EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW, LEVELS REMAIN HIGH
 IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN
 POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX
 AND THORACIC AND ABDOMINAL GANGLIA.
 CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF052754; AAC08536.1; -;
 DR EMBL; AF052753; AAC08536.1; JOINED.
 DR EMBL; AF052754; AAC08535.1; -;
 DR EMBL; AF052752; AAC08535.1; JOINED.
 DR EMBL; K03054; AAA51462.1; -;
 DR EMBL; K03417; AAA51460.1; -;
 DR EMBL; K03416; AAA50965.1; -;
 DR EMBL; K03418; AAA51461.1; -;
 DR EMBL; AF109077; AAD26134.1; -;
 DR EMBL; AF109078; AAD26132.1; -;
 DR EMBL; AF109082; AAD26132.1; JOINED.
 DR EMBL; AF109078; AAD26133.1; -;
 DR EMBL; AF109084; AAD26133.1; JOINED.
 DR EMBL; AF109079; AAD26130.1; -;
 DR EMBL; AF109081; AAD26130.1; JOINED.
 DR EMBL; AF109079; AAD26131.1; -;
 DR EMBL; AF109083; AAD26131.1; JOINED.
 DR EMBL; AF109080; AAD26135.1; -;
 DR EMBL; AE003454; AAF46732.1; -;
 DR EMBL; X02293; CAA26157.1; -;
 DR EMBL; X78920; CAA55523.1; -;
 DR EMBL; X78918; CAA55521.1; -;

DR EMBL; X78919; CAA55522.1; -;
 DR PIR; A00640; GOFPE.
 DR HSP; P11362; IFGK.
 DR FlyBase; FBgn0003731; Egfr.
 DR InterPro; IPR000494; EGFR_L_domain.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recep_L_domain; 2.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00261; FU; 7.
 DR SMART; SM00219; TyrKc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 KW Transmembrane; Glycoprotein; Receptor; Phosphorylation; Transferase;
 KW Tyrosine-protein kinase; ATP-binding; Signal; Alternative splicing;
 KW Developmental protein.
 FT SIGNAL 1 30 POTENTIAL.
 FT CHAIN 31 1426 EPIDERMAL GROWTH FACTOR RECEPTOR.
 FT DOMAIN 31 868 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 869 889 POTENTIAL.
 FT DOMAIN 890 1426 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 938 1198 PROTEIN KINASE.
 FT NP_BIND 944 952 ATP (BY SIMILARITY).
 FT BINDING 971 971 ATP (BY SIMILARITY).
 FT ACT_SITE 1063 1063 BY SIMILARITY.
 FT MOD_RES 902 902 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
 Query Match 28.1%; Score 1931; DB 1; Length 1426;
 Best Local Similarity 32.8%; Pred. No. 1.4e-25; Indels 342; Gaps 42;
 Matches 464; Conservative 180; Mismatches 459;
 QY 37 LRVPKVSASHL-----ELYQCQCVQVQGNLELTLYPT-NASLSFLQDIOEQVGYVLIHQV 91
 DB 109 LSVPSNKEHYHNLRYRNTCTVDGNLKLTLWLPNENLDLSFLDNIREVTGILISHVDV 168
 QY 92 RQVPLQRLRIVGTQLF-----EDNALAVLDNGDPLNNTPTVTGASPGGLRELQLSLT 146
 DB 169 KKVVPFKLQIRGRTLFSLSEBEKALFV-----TYSKMYTLEIFDLR 212
 QY 147 EILKGGVLIQNPQLCVQDITLWKDIFHKNNQLALTIDNRSRACHPCSPMCKSGSCW 206
 DB 213 DVLNGQVGFHNNYLNLCMRITQMSIVNSGTDAYNVDFTAPRECPKCHESCTHG-CWG 271
 QY 207 ESSEDQCSLRTTVCAGGCA--RCKGPLPTDCHEQCAAGCTGPKHSDCLACLFHNSGIC 264
 DB 272 EGPKNCKFSKLTCSQCAGRCYGVKPRECCHLFCAGGCTGPTQKDCIACKNPFDEAVS 331
 QY 265 ELHCPALVTYNTDTFESMPNPEGRYTFGASCTACPNYLYSTDVGSCTLVCPPLHNQEVTA 324
 DB 332 KEBCPPMRKYNPTTYVLETNPEGKYATGATCKVECP--GHLLRDNGACVRSQPDMDKGG 390
 QY 325 EDGTQCEKCSKPCARVCYGLGMEHLREVPVTSANIQEFAGCKKIFGSLAFIPESFDG- 383
 DB 391 E-----CVPNGPCPKTCPGTVLH-----AGNIDSPRNCITVDIGNIRILDQTSGF 437
 QY 384 -DPASNTA-----PLQPEQLQVFETLEITGYLYTISAMPDLSLPDLSVFQNLQVIRGIL 436
 DB 438 QDVYANYTMGPRIPLDPERREVFSTVKETGYLNTIEGTHPQFNLVSFRNLETHGRL 497
 QY 437 HNGAY-SLTLOGLGISWLGRLSRLRELSGSLALIHNNTHLCFVHTVPMQDLFRNPHQALLH 495
 DB 498 MESMFAALAIKSLYSLSLEMLNLKQISSGSWVIQHNRLDLCVSNIRNPAIQKEQKVVV 557
 QY 496 TANRPDECVGEGLACHQLCARGCHVGPGPTQCVCNCSQFLRGQBCVCECRVQLQPLREYV 555
 DB 558 NENLRADLCBKNGTICSDQCNEDGCGWAGTDQCLTKNFNGTCTIADCGVISNAYK--F 615
 QY 556 NARHCLPCHPECPQNSVTCFGPEADQCACVAHYKDPFPCVACRP----- 601

```
Db 616 DNRTCKICHPECR-----TCNGAGADHCQECVHVRDGHQCVSECPKKNKYNDRGVCRECH 669
Qy 602 -----SGVK-----PDLSTMPYWKPF--PDEEG 621
Db 670 ATCDGCTGPKDTIGAGACTTCNLAIINNDATVKRCLLKDCKPD-GY--FWYVHQPQEQ 726
Qy 622 ACQP-----CPI-----NCTH-----632
Db 727 SLKPLAGRAVCRKCHPLCELCCTNYGHEQVCCKTHYKRREQCCECETPCADHYTDEQREC 786
Qy 633 -----SC-----VDLDKG-----CPAQR-----647
Db 787 QRHPECNCGTGGADCKSRNFKLFDANETGYPVNSTWFCNCTSKCPLEMRHNVQYTA 846
Qy 648 -----ASPLTS-----IVSAVVGILLVVVLGVVFGILIKRQOKIRKYT--686
Db 847 IGPYCAASPRSSKITANLDVNMFIITGAVLPTICILCV--TYICRQKQAKETVK 904
Qy 687 MRRLQETELVEPLTPSGAMPNQAMRILKETELRKVKVLGSGAGFYVYKGIWIPDGENV 746
Db 905 MTWALSCEDESEPLRPSNIGANCKLRIVKDAELRKGVLGMAFGRVYKGVVWPEGENV 964
Qy 747 KIPVAIKVLRNTPSPRANKELDEAYVMAGVSPYVSRLLIGICLTSTVQLTQMPYGCCL 806
Db 965 KIPVAIKELLUKSTGAESSEBFLREAYIMASEEHVNLKLLAVCMSSQMLITQLMPLGCL 1024
Qy 807 LDHVRENRLGSGODLLNWCQAKGMSYLEDLVRLVHRDLAARNVLKSPNHVKITDFGL 866
Db 1025 LDVYRNRRDKIGSKALLNWSQTAKGMSYLEEKLVRDLAARNVLVQTESLKIITDFGL 1084
Qy 867 ARLLDIDETEHADGGKVPKIKWMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI 926
Db 1085 AKLLSSDSNEYKAAGKMPKWLALCEIRNRVFTSKSDVWAFVTIWELLTFQGRPHENI 1144
Qy 927 PAREIPDLLEKGERLPQPPCTTDVYMWVKWIMIDSECRPRRELVSFSRWARDPQRF 986
Db 1145 PAKDIPDLIEVGLKLEQPEICSLDIYCTLLSCHWLDAAARPTFKQLTTFVFAEFARDPGRY 1204
Qy 987 VTIONEDLG--PASPLDSTFYRSLLEDD--DMGDLVDAREYLVPOGGFCPPAPGAGG 1041
Db 1205 LAIPGDKFTLPA-----YTQDEKDLRLKLAPTTDGSEAIKAPDDYIQPKAAAPGS- 1256
Qy 1042 MVHHRHSSTRSGGDLTLGLEPSEEAPEP-----RSLAPSEGAGSDVFDG---DLGM 1092
Db 1257 -----HRTDCT-----DEMPKLNRYCKDPKSNKNSSTGDDERDSSAREVG 1296
Qy 1093 GAAKGLQSLTHDPSPLORYSEDPVPLPSETDGYVAPLTCSPQPEYVNPQDVRPOPSP 1152
Db 1297 GNLR-----LDLPVDEDDYLMPTCPQGPNNNNNN-----NP 1328
Qy 1153 REGPLPAARPAAGATLERAKTLPSPKNGVVKDVFAGGAVENPEYL---TFQCGAAPOPH 1208
Db 1329 QNNMAAVGVAAGM-----DLTGVPVSVDNPNPYLLNAQTLGVGESPIPT 1373
Qy 1209 -----PPAFSP-AFDNLYTYMD 1224
Db 1374 QTIGIPWGGPGTMEVKVPMGSEPTSSDHEYIND 1408
```

RESULT 12

```
ERBB_ALV
ID ERBB_ALV
AC P00534;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (BC 2.7.1.112).
GN V-ERBB.
OS Avian leukosis virus.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
ON NCBI_TaxID=11864;
RX [1]
```

```
RP SEQUENCE FROM N.A.
RX MEDLINE=85228222; PubMed=2988784;
RA Nilssen T.W., Maroney P.A., Goodwin R.G., Rottman F.M.,
RA Crittenden L.B., Raines M.A., Kung H.-J.;
RT "c-erbB activation in ALV-induced erythroblastosis: novel RNA
RT processing and promoter insertion result in expression of an
RT amino-truncated EGF receptor.";
RL Cell 41:719-726(1985).
CC -|- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -|- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC IN CHICKS WHICH HAVE BEEN INJECTED WITH THE AVIAN LEUKOSIS VIRUS
CC AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC -|- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC PROTEIN.
CC -|- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M10066; AAA48763.1; ALT_INIT.
CC PIR: A00643; TVCHLV.
CC PIR: B00643; TVFVLV.
CC HSSP: P11362; 1FGK.
CC
CC InterPro: IPR000715; Euk_pkinase.
CC InterPro: IPR001245; Tyr_pkinase.
CC Pfam: PF00069; pkinase; 1
CC PRINTS: PR00109; TYRKINASE.
CC ProDom: PD000001; Euk_pkinase; 1.
CC SMART: SM00219; TyrK; 1.
CC PROSITE: PS00107; PROTEIN_KINASE_TYR; 1.
CC PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
CC PROSITE: PS00111; PROTEIN_KINASE_DOM; 1.
CC Transferrase; Tyrosine-protein kinase; ATP-binding; Oncogene;
CC Glycoprotein; Phosphorylation.
CC DOMAIN 132 399 PROTEIN KINASE.
CC FT NP BIND 138 146 ATP (BY SIMILARITY).
CC FT BINDING 165 165 ATP (BY SIMILARITY).
CC FT ACT SITE 257 257 BY SIMILARITY.
CC SEQUENCE 634 AA; 70891 MW; E705E33A0BE01FCC CRC64;
CC
CC Query Match 25.7%; Score 1749.5; DB 1; Length 634;
CC Best Local Similarity 52.3%; Pred. No. 2.7e-86;
CC Matches 370; Conservative 79; Mismatches 135; Indels 123; Gaps 17;
CC
Qy 587 CAHYKDPPCVACRCSGVKPDLSYMPDWKFPBEGACQPCPNCTHSCVDLDKKGPAEQ 646
Db 3 CAHFIDGPHCVKACPAVLGENDTL-VWKYADANAVCOLCHPNCRCGPGLEGCP---58
Qy 647 RASPLTSIVSAVY-GILLVVVLGVVFGILIKRQOKIRKYTMRLRQLQETELVEPLTPSGA 705
Db 59 NGSKTPSIAAGVVGGLLCLVVGVLGIGLYLRRR-HIVKRLRLRLQERELVEPLTPSGE 117
Qy 706 MPNQAMRILKETELRKVKVLGSGAGFYVYKGIWIPDGENVKIPVAIKVLRNTPSPKANK 765
Db 118 APNQAHLRLKETEPKVKVVLGSGAGFYVYKGIWIPDGENVKIPVAIKVLRNTPSPKANK 177
Qy 766 EILDEAYVMAGVSPYVSRLLIGICLTSTVQLTQMPYGCCLLDHVRENRLGSGODLLNW 825
Db 178 EILDEAYVMAGVSPYVSRLLIGICLTSTVQLTQMPYGCCLLDHVRENRLGSGODLLNW 237
Qy 826 CMOIAGMSYLEDLVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVP 885
Db 238 CVOIAGMSYLEDLVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVP 297
Qy 886 IKWMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPP 945
```



```
Db 298 IKWMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGI PASEISSVLEKGERLPQP 357
Qy 946 ICTIDVTMVMKCMWIDSECRPRRELVSFSEMRDQPFVVIQ-NEDLGPASPLDSTF 1004
Db 358 ICTIDVTMVMKCMWIDSECRPRRELVSFSEMRDQPFVVIQ-NEDLGPASPLDSTF 417
Qy 1005 YRSLEDDDDMGDLVDAEYLVPOQGFPCPDAPGAGGGMVHRRSSSTRSGGDLTLGLE 1064
Db 418 YRTLMEEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
Qy 1065 PSEEEAPRSL-----APSEGAGSDVFDGDLGMAAGLQSLPHTDPSLQRYSEDTVP 1119
Db 450 -----SRTPLLSLSLTSNNSATNCID-----RNGQGHVPVREDSFQRYSSDPTGN 495
Qy 1120 LPSET--DGVVAPLTCSPQPEYVNDVVRPQPSREGPLPAARPAGATLERAKTLPSPGK 1177
Db 496 FLEESIDDDGL-----PAPEYVQ--LMPKKPS----- 524
Qy 1178 NGVVKDVF-----AFGGAIVENPEVLTPOGGAAPQHPHPPAFSPAFDNLXY 1222
Db 527 NQIYNNISLTAISKLPMDSRQNSHSTAVDNPEYL-----NTNQSPLAKTVFESSPY 578
Qy 1223 WDQ-----DPPE-----RGAPPSTFGTPTAENPEYLGLDVP 1254
Db 579 WIQSGNHQINLDNPDYQDFLNETKPNGLLKVPAAENPEYLRVAAP 625

RESULT 13
ERBB AVIER STANDARD; PRT; 604 AA.
AC P00535;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUN-1986 (Rel. 01, Last sequence update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ES4).
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=79685;
RN [1]
RN [2]
RC STRAIN=H;
RX MEDLINE=84026539; PubMed=63113229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erbB gene of avian erythroblastosis virus is a member of the src
RT gene family."
RL Cell 35:71-78(1983).
RN [2]
RX SEQUENCE OF 1-152 FROM N.A.
RA MEDLINE=84223957; PubMed=6328658;
RA Debuire B., Henry C., Benaisa M., Biserte G., Claverie J.-M.,
RA Saule S., Martin P., Stehelin D.;
RT "Sequencing the erbA gene of avian erythroblastosis virus reveals a
RT new type of oncogene."
RL Science 224:1456-1459(1984).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC IN CHICKENS.
CC -1- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
DR EMBL; K02006; AAA42394.1; ALT_INIT.
DR EMBL; K01216; AAA42400.1; -.
DR PIR; A00644; TVYUH.
DR HSSP; P11362; LFGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrcK_1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
KW DOMAIN 132 399 PROTEIN_KINASE.
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
FT CONFLICT 29 29 R -> W (IN REF. 2).
FT CONFLICT 140 140 S -> F (IN REF. 2).
FT CONFLICT 145 145 I -> V (IN REF. 2).
SQ SEQUENCE 604 AA; 67633 MW; 76BCDD06745D609 CRC64;

Query Match 25.0%; Score 1703; DB 1; Length 604;
Best Local Similarity 52.2%; Pred. No. 7.6e-84;
Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 16;

Qy 587 CAHYKPPFCVACPSGVKPDLSYMPIWKEPDEGACQPCINCTHSCVDLDDKGCPAEQ 646
Db 3 CAHFIDGPHCVKACPAVLGENDTL-VRYADANAVCQLCHPNCRTGCGPGLEGCP--- 58
Qy 647 RASPLTSIVSAVV-GILLVVLGVWFGLIKRQKIRKVTMRLLQETELVEPLTPSGA 705
Db 59 NGSKTPSIAAGVVGGLLVVGLGIGLYLRRR-HIVRKTLLRLLQERLEPLTPSGE 117
Qy 706 MPNQAMRILKETELRKVKVLGSGAFGTGVYKGIWIPDGENVKIPVAIKVLRNTPKANK 765
Db 118 APNOAHLRLKETEFKKVKVLGSGAFGTGVYKGIWIPDGENVKIPVAIKVLRNTPKANK 177
Qy 766 EILDEAYVAVGCVSPVSRLLGLCLTSTVQLTQPLPYGCLLDHVRNRRGLSODLLNW 825
Db 178 EILDEAYVAVASVDNPHVCRLGLCLTSTVQLTQPLPYGCLLDYIREHKDNISQYLLNW 237
Qy 826 CMQIAKMSYLEDVRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVP 885
Db 238 CVQIAKGMVLEERLVRHDLAARNLVKTPQHVKITDFGLAKLGADEKEYHAEGKVP 297
Qy 886 IKWMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGI PASEISSVLEKGERLPQP 945
Db 298 IKWMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGI PASEISSVLEKGERLPQP 357
Qy 946 ICTIDVTMVMKCMWIDSECRPRRELVSFSEMRDQPFVVIQ-NEDLGPASPLDSTF 1004
Db 358 ICTIDVTMVMKCMWIDSECRPRRELVSFSEMRDQPFVVIQ-NEDLGPASPLDSTF 417
Qy 1005 YRSLEDDDDMGDLVDAEYLVPOQGFPCPDAPGAGGGMVHRRSSSTRSGGDLTLGLE 1064
Db 418 YRTLMEEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
Qy 1065 PSEEEAPRSL-----APSEGAGSDVFDGDLGMAAGLQSLPHTDPSLQRYSEDTVP 1119
Db 450 -----SRTPLLSLSLTSNNSATNCID-----RNGQGHVPVREDSFQRYSSDPTGN 495
Qy 1120 LPSET--DGVVAPLTCSPQPEYVNDVVRPQPSREGPLPAARPAGATLERAKTLPSPGK 1177
Db 496 FLEESIDDDGL-----PAPEYVQ--LMPKKPS----- 524
Qy 1178 NGVVKDVF-----AFGGAIVENPEVLTPOGGAAPQHPHPPAFSPAFDNLXY 1222
Db 525 --VQNIYNNISLTAISKLPMDSRQNSHSTAVDNPEYL-----NTNQSPLAKTVFE 574
Qy 1219 NLYIYWDQPPERGAFFPTFGTPTAENPEY 1248
```



```
Db 575 SSPYWTQSGNHQ-----INLDNPDY 594
RESULT 14
ERBB AVIEU STANDARD; PRT; 540 AA.
AC P11273;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ts167).
OC Viruses; Retrod viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=103898;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87064459; PubMed=2878364;
RA Choi O.R., Trainor C., Graf T., Beug H., Engel J.D.;
RT "A single amino acid substitution in v-erbB confers a thermolabile
RT phenotype to ts167 avian erythroblastosis virus-transformed erythroid
RT cells.";
RL Mol. Cell. Biol. 6:1751-1759 (1986).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
Db EMBL; M13179; AAA42401.1; --
DR PIR; A25231; TVFVEB.
DR HSP; P11362; IFKG.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
DR PROSITE; PS00011; PROTEIN KINASE DGM; 1.
KW Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399
FT NP BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT SITE 257 257 BY SIMILARITY.
FT VARIANT 270 270 H -> D (IN THERMOLABILE V-ERBB).
SQ SEQUENCE 540 AA; 60412 MW; 5B53297AA068B65D CRC64;
Query Match 23.9%; Score 1630; DB 1; Length 540;
Best Local Similarity 55.1%; Pred. No. 5.2e-80;
Matches 341; Conservative 69; Mismatches 119; Indels 90; Gaps 14;
QY 587 CAHYKDPFPVACPSGVKFDLSVMPITWKPFDDEGACQPCINCTHSCVDLDKGGPABQ 646
Db 3 CAHFIDGPHCVKACPAAGVLGENDTL-VWKYADANAVCQLCHPNCTCKGPGLEGCP-- 58
QY 647 RASPLTSIVSAVV-GILLVVVGVVFGILIKRQOKIRKVTMRLLQETELVPLTPSGA 705
Db 59 NGSKTFSIAGVVGGLLCLVVGIGLYLRRR-HIVRKRTLRLQLQERELVPLTPSGE 117
QY 706 MPNQAOIRILKTELKRVKVLGSGAFCTYKGIWIPDGENVKIPVAIKVRENTSPKANK 765
Db 118 APQAHILRIILKEFEKKVKVLGFGAFCTYKGLWIPGEKVTIPVAIKELRENTSPKANK 177
QY 766 EILDEAYVMAGVSPYVRLIGLICLTSTVOLTPQLMPYGLLDHVRNCRGLSGDQLLNW 825
|||||
Db 178 EILDEAYVMASVDNPHVCRLLGLICLTSTVOLTPQLMPYGLLDYIREHKDNIGSQYLLNW 237
QY 826 CMOIAKMGYLEDVRLVHRDLAARNVLKSPNHNKIDTDFGLARLLDIDETEHADGCKVP 885
Db 238 CVQIAKMGNYLEERHNVHRDLAARNVLVKTPOHVKITDFGLAKQLGADEKEYHAEKGKVP 297
QY 886 IKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPP 945
Db 298 IKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPIASEISSVLEKGERLPQPP 357
QY 946 ICTIDVYIMVVKWIMDSCEPRFRELVSERFEMARDPQRFVVO-NEDLGPASPLDSTF 1004
Db 358 ICTIDVYIMVVKWMSDASRPFRELIAEFESKWARDPPRYLVLIQDERMHLPSPTDSKF 417
QY 1005 YRSLLEDDMDGDLVDAEYLVPOQGFPCDPAPGAGGMVHHRHRSSTRSGGGLTLGLE 1064
Db 418 YRTLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
QY 1065 PSEEAAPRSL-----APSEGAGSDVFDGDLGMAAKGLQSLTHDPSPLQRYSEDTVP 1119
Db 450 -----SRTPLLSSLSATSNNNSATNCIDRNGG-----H----- 476
QY 1120 LPSETDGYVAPLTCSPQPEYVNPQDVVRPQPPSPREGPLPAARPAAGAT-LERAKTLPQKN 1178
Db 477 -PVREDGFL-----PAPEYVNG--LMPKKPSTAMVQNIYINYSITAKLPIDSRVQN 527
QY 1179 GVVKDVFAGFAGVENPEYL 1197
Db 528 -----SHSTAVDNPEYL 539
RESULT 15
EGFR CHICK STANDARD; PRT; 703 AA.
AC P13387;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (BC 2.7.1.112) (CER)
DE (Fragment).
GN EGFR.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88261272; PubMed=3260329;
RA Lax I., Johnson A., Howk R., Sap J., Bellot F., Winkler M.,
RA Ullrich A., Vennstrom B., Schlessinger J., Givol D.;
RT "Chicken epidermal growth factor (EGF) receptor: cDNA cloning,
RT expression in mouse cells, and differential binding of EGF and
RT transforming growth factor alpha.";
RL Mol. Cell. Biol. 8:1970-1978 (1988).
CC -1- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF.
CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC dimerization, internalization of the EGF-receptor complex,
CC induction of the tyrosine kinase activity, stimulation of cell DNA
CC synthesis, and cell proliferation.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:22:34 ; Search time 48.3575 Seconds
(without alignments)
5347.444 Million cell updates/sec

Title: SEQ4-5-25-14
Perfect score: 6812
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYGLDVPV 1255

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues 671580
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp Vertebrate.*
- 14: sp Unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6183	90.8	1259	6 O18735	O18735 canis famil
2	3140	46.1	1209	11 Q9QX70	Q9QX70 rattus norv
3	3109	45.6	1210	11 Q9EP98	Q9EP98 mus musculu
4	2739	40.2	1165	13 Q9YH40	Q9YH40 xiphophorus
5	2729.5	40.1	1137	13 Q9W5F6	Q9W5F6 gallus gall
6	2292	33.6	1328	13 P79754	P79754 fugu rubrip
7	2000.5	29.4	1433	5 Q9BIH9	Q9BIH9 anopheles g
8	1782.5	25.9	419	4 Q9UK79	Q9UK79 homo sapien
9	1739	25.5	367	11 Q8R2X1	Q8R2X1 mus musculu
10	1720	25.2	729	15 Q86712	Q86712 avian rous-
11	1718	25.2	567	15 Q86714	Q86714 avian rous-
12	1697.5	24.9	412	4 Q8WYV0	Q8WYV0 homo sapien
13	1653.5	24.3	962	15 Q64895	Q64895 avian eryth
14	1645	24.1	545	15 Q85468	Q85468 avian eryth
15	1500.5	22.0	655	11 Q9WVF5	Q9WVF5 mus musculu
16	1484.5	21.8	643	11 Q9ERV6	Q9ERV6 mus musculu

17	1275	18.7	1193	5 Q9YIX8	Q9YIX8 ephydatia f
18	1197.5	17.6	1368	5 Q23821	Q23821 caenorhabdi
19	1175	17.2	1717	5 Q26566	Q26566 schistosoma
20	1118	16.4	527	13 Q90836	Q90836 gallus gall
21	995.5	14.6	478	11 Q9FSE0	Q9FSE0 rattus norv
22	934.5	13.7	599	13 Q9FSH2	Q9FSH2 gallus gall
23	906	13.3	165	4 Q14256	Q14256 homo sapien
24	887	13.0	176	11 Q923V5	Q923V5 rattus norv
25	806.5	11.8	346	13 P11776	P11776 xiphophorus
26	778	11.4	435	5 Q8SZW1	Q8SZW1 drosophila
27	756	11.1	1362	13 Q9PVZ4	Q9PVZ4 xenopus lae
28	754.5	11.1	311	13 Q9N162	Q9N162 xiphophorus
29	735.5	10.8	1671	5 Q9NUV5	Q9NUV5 biophalar
30	723	10.6	149	6 Q9BG66	Q9BG66 oryctolagus
31	710.5	10.4	331	4 Q9BUD7	Q9BUD7 homo sapien
32	701.5	10.3	1368	13 Q8UW85	Q8UW85 paralicthy
33	696	10.2	1418	13 Q93457	Q93457 scophthalmu
34	679.5	10.0	1369	13 Q8UW86	Q8UW86 paralicthy
35	670	9.8	1358	13 Q73798	Q73798 xenopus lae
36	659	9.7	1412	13 Q8UW84	Q8UW84 paralicthy
37	657.5	9.7	1472	5 Q9US48	Q9US48 bombyx mori
38	645	9.5	1245	13 Q9YGH8	Q9YGH8 scophthalmu
39	642.5	9.4	1418	13 Q8UW83	Q8UW83 paralicthy
40	634	9.3	1371	11 Q9QVW4	Q9QVW4 rattus sp.
41	626	9.2	2144	5 Q9VD94	Q9VD94 drosophila
42	605	8.9	935	4 Q96L35	Q96L35 homo sapien
43	601	8.8	987	11 Q91YM0	Q91YM0 mus musculu
44	598	8.8	987	11 Q99MR2	Q99MR2 mus musculu
45	587.5	8.6	1036	4 Q07912	Q07912 homo sapien

ALIGNMENTS

RESULT 1

O18735 PRELIMINARY; PRT; 1259 AA.
ID O18735
AC O18735;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Erbb-2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Yokota H.;
RT "CDNA cloning of erbb-2 from canine mammary gland.";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB008451; BAA23127.1; -
DR HSSP; P11362; IFGK.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; YFP_motif.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Pfam; PF02757; YLP_2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN 1.
DR PROSITE; PS0107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1259 AA; 137989 MW; E37364D49C4ACD46 CRC64;

Query Match 90.8%; Score 6183; DB 6; Length 1259;
Best Local Similarity 90.5%; Pred. No. 0;
Matches 1144; Conservative 43; Mismatches 63; Indels 14; Gaps 4;

Qy 1 MELAAACRWGLLLALLPPGAASSTVCTFNNTVSWFLRVKVSASHLE----LYQGCQVV 56
Db 1 MELAAACRWGLLLALLPPGAASSTVCTFNNTVSWFLRVKVSASHLE----LYQGCQVV 56

Qy 57 QGNLELYLPTNASLSFLQDIQEVQYVLIHAHNOVROVPLQRLRIRVGTQLFEDNYALAV 116
Db 57 QGNLELYLPTNASLSFLQDIQEVQYVLIHAHNOVROVPLQRLRIRVGTQLFEDNYALAV 116

Qy 117 LDNGDPLNNTTTPVTGASPGGLREQLRLSLTEILKGGVLIQRNPOLCYQDTILMKDIFHN 176
Db 117 LDNGDPLNNTTTPVTGASPGGLREQLRLSLTEILKGGVLIQRNPOLCYQDTILMKDIFHN 176

Qy 177 NQALALTIDNRSRACHPCSPMKGRCWGBSSSDCQSLRTVTCAGGACRCKGPLPTDCC 236
Db 177 NQALALTIDNRSRACHPCSPMKGRCWGBSSSDCQSLRTVTCAGGACRCKGPLPTDCC 236

Qy 237 HEQCAAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRTYFGASCV 296
Db 237 HEQCAAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRTYFGASCV 296

Qy 297 TACPYNLSTDVGSCITLVCPHNOQVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRV 356
Db 297 TACPYNLSTDVGSCITLVCPHNOQVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRV 356

Qy 357 TSANIQFAGCKITFGLAFIPESFDGDPASNTAPLOEQLOVFTLEETGYLYISAMP 416
Db 357 TSANIQFAGCKITFGLAFIPESFDGDPASNTAPLOEQLOVFTLEETGYLYISAMP 416

Qy 417 DSLPDLVSFQNLQVIRGRILHNGAYSLTLOGLSISWGLSLRGLSGLALIHNTLCP 476
Db 417 DSLPDLVSFQNLQVIRGRILHNGAYSLTLOGLSISWGLSLRGLSGLALIHNTLCP 476

Qy 477 VHTVPDQLFENPHQALLHTANRDECEVGBGLACHQALCHGCHWGCPQTQVNCQSLR 536
Db 477 VHTVPDQLFENPHQALLHTANRDECEVGBGLACHQALCHGCHWGCPQTQVNCQSLR 536

Qy 537 GQECVECRVLQGLPREYVNAHCLPHCEQONGSVTCFPGPADQCVCAHYKDPFFC 596
Db 537 GQECVECRVLQGLPREYVNAHCLPHCEQONGSVTCFPGPADQCVCAHYKDPFFC 596

Qy 597 VARPSPGVKPLDLSYMTWKPEDEGACQPCPINTCHSCVDLDDKCPAQRASPLTIVS 656
Db 597 VARPSPGVKPLDLSYMTWKPEDEGACQPCPINTCHSCVDLDDKCPAQRASPLTIVS 656

Qy 657 AVVGILLVVLGVVFGILIKRRQKIRKYMRLLOETELVEPLTPSGAMPNQAMRLK 716
Db 657 AVVGILLVVLGVVFGILIKRRQKIRKYMRLLOETELVEPLTPSGAMPNQAMRLK 716

Qy 717 ETELKRVKVLGSGAFYVYGIWIPDGENVKIPVAIKVLRNTSPKANKILDEAYVMAG 776
Db 717 ETELKRVKVLGSGAFYVYGIWIPDGENVKIPVAIKVLRNTSPKANKILDEAYVMAG 776

Qy 777 VGSYVSVRLGICLTSTVQLVTLQMPYCYLLDHYRENRGLSGDILNWCQIAKGSYL 836
Db 777 VGSYVSVRLGICLTSTVQLVTLQMPYCYLLDHYRENRGLSGDILNWCQIAKGSYL 836

Qy 837 EDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEYHADGGKVPKMWALLESILR 896
Db 837 EDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEYHADGGKVPKMWALLESILR 896

Qy 897 RRFTHQSDVMSYGVTVMLTFGAKPYDGIPIAREIPDLLEKGERLPQPICTIDVYIMV 956
Db 897 RRFTHQSDVMSYGVTVMLTFGAKPYDGIPIAREIPDLLEKGERLPQPICTIDVYIMV 956

Qy 957 KCMWIDSECRPRFRELVSSEARMARDPQRFVIONEDLGPASPLDSTFYSLLEDDMDGD 1016
Db 957 KCMWIDSECRPRFRELVSSEARMARDPQRFVIONEDLGPASPLDSTFYSLLEDDMDGD 1016

Qy 1017 LVDABEYLVPOQGFPCPDPAFGAGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSPLA 1076
Db 1017 LVDABEYLVPOQGFPCPDPAFGAGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSPLA 1076

Db 1016 LVDABEYLVPOQGFPCPDPAFGAGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSPLA 1075
Qy 1077 PSEGAGSDVFDGDLGMAAGKLOSLTHDPSPLQRYSEDTVPLPSETDGYVAPLTCSPQ 1136
Db 1076 PSEGAGSDVFDGDLGMAAGKLOSLTHDPSPLQRYSEDTVPLPSETDGYVAPLTCSPQ 1135
Qy 1137 PEVNOPDVVRPQPPSPREGPLPAARPAAGATLER-----AKTLPSPGKNGVVVKDVFAFGAV 1191
Db 1136 PEVNOPDVVRPQPPSPREGPLPAARPAAGATLER-----AKTLPSPGKNGVVVKDVFAFGAV 1195
Qy 1192 ENPEYLTPOGGAAPQHPPPAFSPAFDNLYYWDQDPPERGAPESTFKGTPTAENPEYLG 1251
Db 1196 ENPEYLTPOGGAAPQHPPPAFSPAFDNLYYWDQDPPERGAPESTFKGTPTAENPEYLG 1255
Qy 1252 DVPV 1255
Db 1256 DVPV 1259

RESULT 2
Q9QX70 PRELIMINARY; PRT; 1209 AA.
AC Q9QX70;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Epidermal growth factor receptor.
GN EGFR.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=101116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FISHER; TISSUE=LIVER;
RX MEDLINE=90258888; PubMed=2342466;
RA Petch L.A., Harris J., Raymond V.W., Blasband A.J., Lee D.C.,
RA Earp H.S.;
RT "A truncated, secreted form of the epidermal growth factor receptor is
RT encoded by an alternatively spliced transcript in normal rat tissue.";
RL Mol. Cell. Biol. 10:2973-2982(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FISHER; TISSUE=LIVER;
RA Petch L.A.;
RL Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=FISHER; TISSUE=LIVER;
RA Guttridge K., Dawson T.L., Earp H.S.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; M37394; AAF14008.1; --
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00319; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00011; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
KW ATP-binding; Receptor; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1209 AA; 134891 MW; 96FEE7F6CC1B773 CRC64;

Query Match 46.1%; Score 3140; DB 11; Length 1209;
Best Local Similarity 49.9%; Pred. No. 1e-226;

QY 408 GYLIVISAWPSLPLSLVQFQNLQVIRGRILHNGAVSLTLQGLGSLWGLRLSLRELGSGLAL 467
Db 391 DILNIQWPKELNDLSFSSLTITQGRSLFKRFSLMVMRPTLTSLGLRLSLREISDGSVY 450
QY 468 IHNHNLCLFVHTVPDOLFRNPH-QALLHTANRDECEVGEGLACHOLCARGHCWGPPT 526
Db 451 ISQNAHLCYHTVNTQTLFRGSRVANSLSNRPMACVADGRVCDPLCSDSCGCGPDP 510
QY 527 QCVNCSOFLRGQECVEECRVLQGLPREYVNAH-CLPCHPECOQNGSVTCFQPEADQCV 585
Db 511 QCLSCRYSRHGTCAVCHFNISGPIREFAGLNGVCVACHPECKPQTCKASCTGFGADECM 570
QY 586 ACAHYKDPPECVACRCPGVDPDLSYMPIWKFPEDEGACQCPNCTHSCVDLDDKGPAP 645
Db 571 ACTFRDGPCTCMSCPAGVN-DGEKGLIFKFPNREHCEPCHQCTQCGSGPGLNDC--- 626
QY 646 QRASPLTSIVSAVVGILLVVLGVVF-----GILIKRQKQKIRKVTMRLLQETELVE 698
Db 627 LEAARLTISSQITGIALGVPAGLIFCLVLFLOGLVHRGLAIRKRAMRYLESGESFE 686
QY 699 PLTPSGAMPNQAQMRILKETELRKVKVLGSGAFQVYKGIWIPDGENVKIPVAIKVLREN 758
Db 687 PLGP-GEKGTKVHARILKPSDLRKIKPLGSGVFGTVSGKFWIPEGETVKIPVAIKTIQDS 745
QY 759 TSPKANKEILDEAYMAGVSPVSVRLGLICLTSTVOLVTLQMPYGCCLLDHVRNRLG 818
Db 746 SGRQTFTETDHLSSMGLDHPYIVRLGICPGTCLQVTLQSSHGSLLEHRIHQHTSLD 805
QY 819 SODLLNMCMOIAKMSVLEDRVLVHRDLAARNLVKSPNKHVITDFGLARLLDIDEYH 878
Db 806 PORLLNMCVQIAKMYLEBHRVHKNLAARNILLKNDYQVQISDYGVADLLYPDDKKYV 865
QY 879 ADGKVPKIMWALESILRRFTHQSDVMSYGVTVWELMTFCAPYDGIPIAREIPDLLEK 938
Db 866 YSETKTPKIMWALESILFRYTHQSDVMSYGVTVWEMWFGAPYASVQVQEPVSVLEK 925
QY 939 ERLPQPPICITDVYIMVWKMDISECRPRELVSEFSRMDRQRFVVIQNEIDLGPAS 998
Db 926 ERLSQAICITDVYIMVWKMDISECRPRELVSEFSRMDRQRFVVIQNEIDLGPAS 980
QY 999 PLDSTFVRSLEDDMDGLVDAEYLVPOQGFPCDPAPGAGGVMVHRHSSSTRSGGD 1058
Db 981 -----EUGMGFEFL-----RGSER-----GL 997
QY 1059 LTLGLEPSEBAPRSLAPSEGAGSDVFDGLMG-----AAKGLQSLPTHDPSPLOQ----- 1110
Db 998 LEADLEDEE-----GLGDRFATPSLOPSPSWTSFSQINSYM 1036
QY 1111 -----RVSEDPVLPSETGYVAPLTCSPQ- EYVNO-----PDVRPQP 1150
Db 1037 VMTQLRYD-----FAVSOGHIGYLPMSPPVDITRQLWYQSRSLSSVRTLPRDSAFRR 1090
QY 1151 SPREGPL--PAARPAGATLERAKTSLPGKGVVVDVAFGAVENPEYLTPOGGAPOPH 1208
Db 1091 SSREAECEDEGAOCAGIFRVR-----FGSERGN-----POGG----- 1122
QY 1209 PPAPFAPFNLYWDDPPERGAPPSTFKGTPTAENPE 1247
Db 1123 -----QORKLSTASSPSSFKTWADEDE 1146

RESULT 7

Q9BIH9
ID Q9BIH9 PRELIMINARY; PRT; 1433 AA.
AC Q9BIH9
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative epidermal growth factor receptor (Fragment).
GN EGFR.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;

OC Anopheles.
OX NCBI_TaxID=7165;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SU4;
RA Lycett G.J.;
RT Cloning, expression and localisation of the Anopheles gambiae
RT epidermal growth factor receptor.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL; AJ301655; CAC35008.1; --
DR HSP; P11362; 1FGK.
DR InterPro; IPR000345; CytoC_heme_bind.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser thr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 7.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN 4.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00111; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Receptor..
FT NON_TER 1
SQ SEQUENCE 1433 AA; 159585 MW; E3D9D88967724F07 CRC64;

Query Match 29.4%; Score 2000.5; DB 5; Length 1433;
Best Local Similarity 32.6%; Pred. No. 2.8e-142;
Matches 464; Conservative 191; Mismatches 383; Indels 385; Gaps 36;

QY 50 YGCGVQVQGNLEILYLTPTNASISFLQDIOEVQGYVLIAHNQVRQVPLQRLIRVGRQTOLF- 108
Db 25 YTNCTVVDGNLEITWITQNTIDNLNLFQIHREVTGYVLISLYDLQVILPRLQIIRGRTFK 84
QY 109 ----ENYALAVLDNGDPLNNTPTVTGASPGGLRELQRLSLTEILKGVILIQNPOLCYQ 164
Db 85 LNKWEAYGLFV-----SFSHMTLELPALRDILGSGVGFNNYLNCHM 128
QY 165 DTLLWKDI-FHKNNQLALTIDNRSRACHPCSPCKSGSCWGESSEDCQSLTRTYCAGG 223
Db 129 KSNWEEILLAPQTSQYTFNFSSPVPVCPCHPCEVG-CWGEHAHNCQRFSLKNCSPQ 187
QY 224 CA--RCKGELPTDCHEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFES 281
Db 188 CSQGRCFGPKPRECCHLFCAGGCTGPTQSDCLACKNFYDDGVCKQECPPMQIYNPTNFW 247
QY 282 MNPEGRYTFGASCTVACPYNYLSTDVGSCTLVCPILHNOEVTADGTQRCCKCKPCARV 341
Db 248 EPNPDGKYAGATCVKACP-EHLLKDNAGCAVRCKPKGMPQNSE-----CVPCKGVCPKT 301
QY 342 CYCLGNHLEHREVRVTSANIQFAGCKKIPGSLAFLPESFGDPSANT-----APLQ 393
Db 302 CPGEGLVH-----SDNIGNYKDCITIEGSLLEILDQSPDGFQVYVNFSPGRYIKID 353
QY 394 PEOLQVFETLEITGYLYISAWPDSLPDLSVFONLQVIRGRIILHNGAY-SLTQGLGISW 452
Db 354 PDELEVFSTVKBITGFINQAHHPNPTTLNFRNLEVGVRQLKENLFSVIVTKSLKS 413
QY 453 LGLRSLRELGSGLALIHNNHLCFVHTVPDQWLFNRPHQALLHTANRPEDECVGEGLACH 512
Db 414 LELSLKRVNSGSIIVLENSDLCFVEDIDWSLEKKSSDHEVMVQKRNATECHEEGWEC 473
QY 513 QLCARGHCWGPPTQCVNCSQFLRGQECVBECEVRVLOGLPREY-VNARHCLPCHPECPQN 571
Db 474 EQCSKAGCWGKPEQCLECKNVKYGKCLDSCK---SLPRLYSVDSKTCGDCHQECCKD-- 528

```
QY 572 GSVTCFGPEADQCACAHYKDPFPCVACRP----- 601
Db 529 --FCYGNEDCGSCMNVKDGRCFVACPTTKHAMNGTCINCHKTCVCGRGRDPTIAPD 585
QY 602 -----SGVKPDLUSYMPIWKFPD-----BEG----- 621
Db 586 GCISCDKAIIGSDAKIERLKMDESCPDGYYSYDVLYQEGPLKQLSGKAVCRKCHPRCK 645
QY 622 -----ACOPCPINCT-----HSC 634
Db 546 CTGYGFHQFQCECTGYKKGQCEDECPQDPYANEETRICUPCHQECRGCHGLDDHHEC 705
QY 635 VDL-----DD-----KGCPAEQ-----RASPLTSI 654
Db 706 RNLKLFEGDPYDNATTFTCVSNCPASHPYKRFQPEAGKIGYCSADSMQSGRLRTEPQTQV 765
QY 555 VSAVVGILLVVVLGVVFGI-----LTKRQOKIRKYTMRLRLQETELVPLPTSPGAMPNOAQ 711
Db 766 KIVMGSMVALILLVCVFGIAFVFSRHNKDKDAVMTMALAGCEDSEPLRPSNVGPNLT 825
QY 712 MRILKETELRKVKVLGSGAFCTVYKGIWIPGENVKIPVAIKVLRENTSPKANKEIILDEA 771
Db 826 LRITKEAIRGGVGLMGAFGRVFGVMWPGESVKIPVAIKVLWMSGSESSKEFELEA 885
QY 772 YWAGVGSPPYVSRILIGICTSTVOLVTLMPYGCILLDHVRNENRGLSGQDILNMCQIAK 831
Db 886 YIMASVEHPNLLKLLAVCMTSQMMLITQLMPLGCLLDYVRNNKDKIGSKALLNWSQIAR 945
QY 832 GMSYLEVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETFYHADGKVPKIMKAL 891
Db 946 GMAYLEERLVRDLAARNVLVOTPSVKITVFGIAKLLDPDSDEYRAAGKMPDKMAL 1005
QY 892 ESILRRRTHOSDVMSGVTVWELMTGAKYDGIIPAREIPDLLEKGERLPPOPICTIDV 951
Db 1006 ECIHRVFTSKSDVWAFGITTWELTYGARDYENVPKDVPELIEIGHKUPQPDICSLDV 1065
QY 952 YIMVVKCWMIDSECRPRELVSFBSFMRARDPQRFVVIQNEBDLGPASPLDSTFYRSLED 1011
Db 1066 YCILLSCVLDADARPTFKLAETFAEKARDPGYLM-----PGDKFEMELPSVTN 1116
QY 1012 DDMGDLV-----DAEYLVVPOQGFPCDDPAPGAGGM 1042
Db 1117 QDEKDLIRTLAPVAMAAAAAGASNVDPSTIAETDEYLOPKTRPSIMLPGPSA--- 1173
QY 1043 VHRHRSSTSGGDLTLGLEPSEERAPRS-----PLAP---SEGAGSDVPDGLGGM 1093
Db 1174 -----VEPS-DEMPKSLRYCKOPLKDDDDTDGKGKEV-----GVG 1207
QY 1094 AAKGLQSLPHTDPSFLQRYSDPTVPLPSETDGYVAPLTCSPQPEYVYNQDVRPQPPSPR 1153
Db 1208 GIR-----LNLPLDEDDYIMP-TCOSQ---NQS----- 1231
QY 1154 EGPLPAAPAGATLERAKTUSPGKNGVVKOVFAFGGAVENDEYL-----TPQGGAA 1204
Db 1232 -----TPG-----YMDLIGVPASVDNPEYLMGSTQAIAGLAQSGMG 1267
QY 1205 PQHPHPPAFSPADNLYWDDQPPERGAPPSTFKGTPTAENPE 1247
Db 1268 --PHTPP-----PPNTPNGMPTHQHSQ 1287

RESULT 8
QY 09UK79 PRELIMINARY; PRT; 419 AA.
AC Q9UK79;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DR 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
GN HER-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
OX NCBI_TaxID=9606;
RN [1]
RX SEQUENCE FROM N.A.
MEDLINE=99415951; PubMed=10485918;
Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;
"The HER-2/neu receptor tyrosine kinase gene encodes a secreted
autoinhibitor.";
Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874 (1999).
RN [2]
RX SEQUENCE FROM N.A.
Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;
Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
EMBL; AF177761; AAD56009.2; -
InterPro; IPR000494; EGFR_L domain.
InterPro; IPR002174; Furin-Like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L domain; 1.
DR SMART; SM00261; FU; 1.
SQ SEQUENCE 419 AA; 45472 MW; FEC1BE347E2D030C CRC64;

Query Match 25.9%; Score 1762.5; DB 4; Length 419;
Best Local Similarity 84.7%; Pred. No. 4.7e-125;
Matches 337; Conservative 9; Mismatches 37; Indels 15; Gaps 4;

QY 1 MELAALCRWGLLLALLPPGAASQTQCTFNNFTVSFVLRVPKVSASHLE-----LYQGQCVV 56
Db 1 MELAALCRWGLLLALLPPGAASQTQCTGDMK-----LRLPASPEHLDMRLRLHYQGQCVV 56
QY 57 QGNLEUTYLPNTASFLQDIOEVQGVLIHNVQVPLQRLRIVRGTLQFEDNYALAV 116
Db 57 QGNLEUTYLPNTASFLQDIOEVQGVLIHNVQVPLQRLRIVRGTLQFEDNYALAV 116
QY 117 LDNGDPLNNTPTVYGASPGRLQLRSLEILKGGVLTORNPOLCYQDITLWKDIFHN 176
Db 117 LDNGDPLNNTPTVYGASPGRLQLRSLEILKGGVLTORNPOLCYQDITLWKDIFHN 176
QY 177 NQALALITDNRSRACHPCSPCKGSRGWESSEDCSLTRTVTCAGGCARCKPLPTDCC 236
Db 177 NQALALITDNRSRACHPCSPCKGSRGWESSEDCSLTRTVTCAGGCARCKPLPTDCC 236
QY 237 HEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCV 296
Db 237 HEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCV 296
QY 297 TACPYNVLSVDVSCSTLVCPLNHNOETAEQGTORCEKSKPCARVCYGLGMEHLREVRAV 356
Db 297 TACPYNVLSVDVSCSTLVCPLNHNOETAEQGTORCEKSKPCARVCYGLGMEHLREVRAV 356
QY 357 TSANIOEFAGCKIFGSLAPLPSFDGDPASNTAPLQP 394
Db 352 VPVPLRMQPG--PAHPVLSFLRPSWDLVSFAFYSLPLAP 387

RESULT 9
QY 08R2X1 PRELIMINARY; PRT; 367 AA.
AC Q8R2X1;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DR 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RX SEQUENCE FROM N.A.
Strausberg R.;
Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
EMBL; BC027080; AAH27080.1; -
KW Hypothetical protein.
SQ SEQUENCE 367 AA; 40163 MW; 0BE03395F9E101B0 CRC64;
```

Query Match 25.5%; Score 1739; DB 11; Length 367;
Best Local Similarity 88.0%; Pred. No. 2.3e-123;
Matches 323; Conservative 11; Mismatches 33; Indels 0; Gaps 0;

QY 889 MALESILRRRFTTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPPOPICT 948
DB 1 MALESILRRRFTTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPPOPICT 60

QY 949 IDVYIMVVKCWMIDSECRPRFRELVSFMRMARDPQRFVVIQNEIDLGPSPMDSTFYRSL 1008
DB 61 IDVYIMVVKCWMIDSECRPRFRELVSFMRMARDPQRFVVIQNEIDLGPSPMDSTFYRSL 120

QY 1009 LEDDDMGDLVDAREYLVPOQGFPCDPAPGAGGMVHRHSSSTRSGGDLTLGLPSEE 1068
DB 121 LEDDDMGDLVDAREYLVPOQGFSPDPALGTGTAHRHSSSARSGGGLTLGLPSEE 180

QY 1069 EAPRSLAPSEGAGSDVFDGDLGMAAGKGLQSLPHTDPSLPQRYSEDPTVPLPSETDGYV 1128
DB 181 EAPRSLAPSEGAGSDVFDGDLVGVTKGLQSLSPHDLPLQRYSEDPTVPLPSETDGYV 240

QY 1129 APLTCSQPQYVNVQDVVRPQSPREGPLPAARPAGATLERAKTLSPGKGVVVDVFAFG 1188
DB 241 APLTCSQPQYVNVQDVVRPQSPREGPLPAARPAGATLERAKTLSPGKGVVVDVFAFG 300

QY 1189 GAVENPEYLTPOGGAAPQHPFPAFPAFONLYWDOQPPERGAPSTFGTPTAENPEY 1248
DB 301 GAVENPEYLTPOGGAAPQHPFPAFPAFONLYWDOQPPERGAPSTFGTPTAENPEY 360

QY 1249 LGLDVPV 1255
DB 361 LGLDVPV 367

RESULT 10

Q86712 ID Q86712 PRELIMINARY; PRT; 729 AA.
AC Q86712;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Polyprotein.
DE POLYPROTEIN.
GN POLYPROTEIN.
OC Avian rous-associated virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60725.1; -.
DR HSSP; P03322; 1A6S.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR004028; Retro M.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF02813; Retro M; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00111; PROTEIN KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
DR ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 729 AA; 80649 MW; 84D2F6914EFE1D63 CRC64;

Query Match 25.2%; Score 1720; DB 15; Length 729;
Best Local Similarity 54.8%; Pred. No. 1.8e-121;
Matches 358; Conservative 75; Mismatches 118; Indels 102; Gaps 15;

QY 569 PQNGSVTCFGPBRADQCVCAHYKDPFCVACRCPGVKPDLSYMPIWKFPDEBAGCQPCPI 628
DB 141 PBEATPKTG2--DHCMKCAHFTDGHPCVKACPAVGLGENDTL-VNKYADANAVCOLCHP 197

QY 629 NCHTSCVDLDDKCPAEQASPLTISAVV--GILLVVVLGVVFGILLIKRQOKIRKTYM 687
DB 198 NCTRGCCKGEGECP---NGSKTPSFAAGVVGGLCLVVGIGIGLYLRRR-HIVRRRTL 253

QY 688 RRLQLQTELV3PLTPSGAMPNQAMRILKETELRKVKVLGSGAFGTGVYKGIWIPDGENVK 747
DB 254 RRLQLQRELV3PLTPSGEAPNQAHRLILKETEKKVKVLGSGAFGTGVYKGLWIPGEKVK 313

QY 748 IPVAIKVLRENTSPKANKILDEAYVMAGVGSPPYVSRLLGICLTSTVQLVTLMPYGCCLL 807
DB 314 IPVAIKELREATSPKANKILDEAYVMASVDNPRVCRLLGICLTSTVQLITLMPYGCCLL 373

QY 808 DHVRENRRGLSGQDLLNMCWQIAKMGSYLEDVLRVHRDLAARNVLKSNHVKITDFGLA 867
DB 374 DYIREHKDNI3GQYLLNWCQIAKGNVLEERLVRDLAARNVLKTPHVKITDFGLA 433

QY 868 RLDDIDETBYHADGGKVPKIMMALESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGP 927
DB 434 KLLGADKEKYHAEGGKVPKIMMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGP 493

QY 928 AREIPDLLK3ERLPOPICTIDVYIMVVKCWMIDSECRPRFRELVSFMRMARDPQRFV 987
DB 494 ASEISSVLEK3ERLPOPICTIDVYIMVVKCWMIDADSRPKRELIAEFSKMARDPPRYL 553

QY 988 VTQ-NEDLGPASPLDSTFYRSLLEDDMGDLVDAREYLVPOQGFPCDPAPGAGGMVHR 1046
DB 554 VIQGDREMLP3PTDSKIFYRTLMEEDMEDIVDAEYLVPHQGF----- 598

QY 1047 HRSSSTRSGG3DLTLGLEPSEEEAPRSL-----APSEGAGSDVFDGDLGMAAGKGLQSL 1101
DB 599 -NSPST-----SRTPLLSLSATSNNSATNCID-----RNGQH 631

QY 1102 PTHDPSLPQRYSEDPTVPLPSET--DGYVAPLTCSPQYVNVQDVVRPQSPREGPLPA 1159
DB 632 PVREDSFVQRYSSDPTGNFLESIIDGDL-----PAPEYVNVQ--LMPKKPS----- 675

QY 1160 ARPAGATLERAKTLSPKGVVVDV-----AFGGAVENPEYL 1197
DB 676 -----TAMVQNYNNISLTAISKLPMDSRVYQNSHSTAVDNPYL 715

RESULT 11
Q86714 ID Q86714 PRELIMINARY; PRT; 567 AA.
AC Q86714;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE V-erbB protein (Fragment).
DE V-ERBB.
GN Avian rous-associated virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60727.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.

```
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00111; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Tyrosine-protein kinase.
FT NON TER 1
SQ SEQUENCE 567 AA; 63390 MW; C6D9CBA7ADF725B1 CRC64;

Query Match 25.28; Score 1718; DB 15; Length 567;
Best Local Similarity 55.44; Pred. No. 1.7e-121;
Matches 357; Conservative 73; Mismatches 112; Indels 102; Gaps 15;

QY 578 GPEADQCACAHYKDPFCVAPCGSVKPDLSYMPIWKFPDEEGACQCPINCTHSCVDL 637
DB 1 GP--DHCWKCAHFIDGHCVKACPAAGVLENDTL-VWKYADANAVCQLCHPCNCTRGCKP 57

QY 638 DDKGCPAEQASPLTSTVSAVV-GILLVVVLGVVFGILIKRQKTRKYMYRLLQTEL 696
DB 58 GLEGCP---NGSKTPSTIAAGVVGGLCLVVVGLGIGLYLRRR-HIVRKRTLRLQLREL 113

QY 697 VEPLTPSGAMPNQAEILKTELARKVYLGSGAGFTVYKGIWIPDCGENVKIPVAIKVL 756
DB 114 VEPLTPSGEAPNQAEILKTEPKVKVILGSGAGFTVYKGIWIPDCGENVKIPVAIKEL 173

QY 757 ENTSPKANKELIDAYVMAGVGSYVSRLLGICLTSTVQLVTLQMPYGCLLDHVRENKR 816
DB 174 ENTSPKANKELIDAYVMSVDNPRVCELLGICLTSTVQLITQMLPYGCLLDYIREHKN 233

QY 817 LGSOLLNWCQIAKGSYLEDLVRLVHRDLAARNVLKSNHVKITDFGLARLLIDETE 876
DB 234 IGSQYLLNWCQIAKGMNYLEERLVRDLAARNVLKTPQHVKITDFGLAKLLGADEKE 293

QY 877 YHAGGGKVPKWALESILRRFTHQSDVMSYGVTVWELMTFGAKPDGIPAREIPDLE 936
DB 294 YHAGGGKVPKWALESILHRIYTHQSDVMSYGVTVWELMTFGSKPDGIPASEISVLE 353

QY 937 KGERLPOPPICTIDVYIMVKCWMIDSECRPRFRELSEFSRMARDPQRFVVIQ-NEDLG 995
DB 354 KGERLPOPPICTIDVYIMVKCWMIDADSRKRELIAEFSKWARDPPRVLVQDERMH 413

QY 996 PASPLDSTFRSLDDMDGLVDAEYLVPOQGFCCPDAPAGAGMVHRRHSSSTRSG 1055
DB 414 LPSPTDSKFRYTLMEEDMEDIVDAEYLVPHQGF-----NSPST---454

QY 1056 GGDITLGLSEEEAPSP-----APSEGAGSDVFDGLGMAKGLQSLPTHDPSPLO 1110
DB 455 -----SRTPLLSLSATSNNSATNCID-----RNGQGHVPREDSFVQ 491

QY 1111 RYSDPTVPLPSET--DGYVAPLTCSQPQPEYVQPDVVRPPSPREGPLPAARPAGATLE 1168
DB 492 RYSDPTGNFLESIDGFL-----PAPEVQVQ--LMPKKPS-----526

QY 1169 RAKTLPQKNGVVKDVF-----AFGGAVENPEYL 1197
DB 527 ----TAMVQNGIYNNISLTAISKLPMDSRQYNSHSTAVDNPYL 566

RESULT 12
Q8WYV0 PRELIMINARY; PRT; 412 AA.
AC Q8WYV0;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 44.7 Kda protein.
GN PF3659.
OS Homo sapiens (Human).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ON NCBI_TaxID=9606;
RX [1]
RP SEQUENCE FROM N.A.
RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
```

```
RA Wan D.F., Gu J.R.;
RT "Novel human cDNA clones with function of inhibiting cancer cell growth."
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF318349; AAL55856.1; -.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD00001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00118; EF_HAND; UNKNOWN 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
KW Hypothetical protein.
SQ SEQUENCE 412 AA; 44702 MW; 034397FF3F27D2BC CRC64;

Query Match 24.9%; Score 1697.5; DB 4; Length 412;
Best Local Similarity 80.5%; Pred. No. 3.9e-120;
Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4;

QY 889 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPDGIPAREIPDLEKGERLPQPICT 948
DB 1 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPDGIPAREIPDLEKGERLPQPICT 60

QY 949 IDVYIMVKCWMIDSECRPRFRELSEFSRMARDPQRFVVIQNEDELGPASPLDSTFVRS 1008
DB 61 IDVYIMVKCWMIDSECRPRFRELSEFSRMARDPQRFVVIQNEDELGPASPLDSTFVRS 120

QY 1009 LEDDDMDGLVDAEYLVPOQGFCCPDAPAGAGMVHRRHSSSTRSGGDLTLGLPESEE 1068
DB 121 LEDDDMDGLVDAEYLVPOQGFCCPDAPAGAGMVHRRHSSSTRSGGDLTLGLPESEE 180

QY 1069 EAPRSLPASEGAGSDVFDGLGMAKGLQSLPTHDPSPLOQRYSDPTVPLPSETDGYV 1128
DB 181 EAPRSLPASEGAGSDVFDGLGMAKGLQSLPTHDPSPLOQRYSDPTVPLPSETDGYV 240

QY 1129 APLTCSQPQPEYVQPDVVRPPSPREGPLPAARPAGATLERAKTLPQKNGVVKDVFAG 1188
DB 241 APLTCSQPQPEYVQPDVVRPPSPREGPLPAARPAGATLERAKTLPQKNGVVKDVFAG 300

QY 1189 GAVENPEYLTPOGGAAPQP-----HPPPA---FSPAFDNL 1220
DB 301 GAVENPEYLTPOGGAALPTLLPSAQPSITGTRTHQSGGLHAPSKGHLRQRTQST 360

QY 1221 YWVD-QDPFER-----GAPPSFTFGTPTAEN 1245
DB 361 WWTQCEPEGVRRSPDVSSGREGITSAGIKRWEGPPTTSRGTCARN 410

RESULT 13
Q64895 PRELIMINARY; PRT; 962 AA.
AC Q64895;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Gag, v-erb-A, v-erb-B protein.
GN GAG, V-ERB-A, V-ERB-B.
OS Avian erythroblastosis virus.
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=11861;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9020603; PubMed=1969616;
RA Bruskina A., Jackson J., Bishop J.M., McCarley D.J., Schatzman R.C.;
RT "Six amino acids from the retroviral gene gag greatly enhance the transforming potential of the oncogene v-erb-B."
RT Oncogene 5:15-24(1990).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL; X52209; CAA36459.1; -.
```

```
DR EMBL; X52211; CAA36459.1; JOINED.
DR HSP; P10828; 2NLL.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR000536; Hormone_rec_lig.
DR InterPro; IPR001723; Steroid_receptor.
DR InterPro; IPR001245; Tyr_kinase.
DR InterPro; IPR001628; Znf_C4steroid.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00398; STRDHORMONER.
DR PRINTS; PR00047; STROIDFINGER.
DR ProDom; PD000001; Euk_pkinase; 1.
DR ProDom; PD000035; Znf_C4steroid; 1.
DR SMART; SM00430; HOLI; 1.
DR SMART; SM00219; TyfK; 1.
DR SMART; SM00399; Znf_C4; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; DNA-binding; Nuclear protein; Receptor;
KW Transcription regulation; Transferase; Tyrosine-protein kinase;
KW Zinc-finger
SQ SEQUENCE 962 AA; 108320 MW; 3CSAED791E495CE CRC64;

Query Match 24.3%; Score 1653.5; DB 15; Length 962;
Best Local Similarity 51.7%; Pred. No. 2.9e-116;
Matches 358; Conservative 73; Mismatches 142; Indels 119; Gaps 18;

QY 541 VECRVQLGLPRE-VYNAR-HCLP-----CHPEQC 568
DB 354 IERCOESYLLAFEHYNYRHNIPHEWSKLLMKVADLRMTGAYHASRFLHMKVECPTELS 413
QY 569 PONGSVTCFGEADQCACAHYKDPFCVACRCPGVKPDLSYPIWKFPDEEGACQPCPI 628
DB 414 PQE-----VGP---DHCWKCAHFIDGPHCVKACPAVLGENDTL-VWKYADANAVCQLCHP 465
QY 629 NCHTSCVDLDDKCPAEQRASPLTSIVSAVV-GILLVVVLGVVFGILIKRQOKIRKYTM 687
DB 466 NCTRGCKGPGLEGCP---NGSKTPSIAAGVVGGLCLVVVGLGIGLYLRR-HIVRKRTL 521
QY 688 RLLQETELVEPLTPSGAMPNQAMRLKETELRKVKVLGSGAGFTVYKGIWIIPDGENVK 747
DB 522 RLLQERELVEPLTPSGEAPNQAHRLKETEFKKVKVLGFGAGFTVYKGLWIPEGEKVT 581
QY 748 IPVAIKVLRNTPSKANKEILDEAYVMAGVSPVSRLLGICLTSTVOLVTLQMPYGCCL 807
DB 582 IPVAIKELRATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVOLVTLQMPYGCCL 641
QY 808 DHVRENRLGSLDNLNWCQIAKGMYSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLA 867
DB 642 DYIREHKDNGISQYLLNWCQIAKGMNLYLEERHVRDLAARNVLKTPQHVKITDFGLA 701
QY 868 RLLDIDETEYHAGGKVPKMALESILRRFTHQSDVWSYGVTVVWELMTFGAKPYDGP 927
DB 702 KQLGADEKEYHAGGKVPKMALESILHRYTHQSDVWSYGVTVVWELMTFGSKPYDGP 761
QY 928 AREIPDLLEGERLPQPPICITIDVYIMVKCMWIDSECRPRFRELSEFSRMARDPQRFV 987
DB 762 ASSEISVLEGERLPQPPICITIDVYIMVKCMWGDASRPKPRFRELIAEFSKWARDPPRYL 821
QY 988 VIO-NEDLGSPASPLDSTFYRSLLEDDDMGLVDAEYLVPOQGFCDPAPGAGGMVHR 1046
DB 822 VIQDERMHLPSPTDSKFYRTLMEEDMEDIVDAEYLVPHQGF-----866
QY 1047 HRSSTRSGGDLTLGLEPSEEBEAPRSPAPSGAGSDVFDGLGMAKAGLSLTHDP 1106
DB 867 -NSPST-----SRTPLLSLSLTSN-----NSATKCIDRNGCH-- 898
QY 1107 SPLQRVSEPTVPLPSETDGYVAPLTCSPQPEYVNOQDVRPQPPSPREGPLPAARFAGAT 1166
DB 899 -----PVREBDGL-----PAPEYVNO--LMPKKPSTAMVQNIYINISLT 936
```

```
QY 1167 -LERAKTSLSPGKNGVVKDVFAFGGAVENPEYL 1197
DB 937 AISKLPMDSRYN-----SHSTAVDNPEYL 961

RESULT 14
Q85468 PRELIMINARY; PRT; 545 AA.
ID Q85468;
AC Q85468;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Avian Erythroblastosis virus (Ts34) v-erbB gene.
OS Avian erythroblastosis virus.
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=11861;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=88217326; PubMed=2897102;
RP Scotting P., Vennstrom B., Jansen M., Graf T., Beug H., Haymann M.J.;
RT "Common site of mutation in the erbB gene of avian erythroblastosis
RT virus mutants that are temperature sensitive for transformation.";
RL Oncogene Res. 1:285-278(1987).
DR EMBL; X06943; CAA30024.1; -.
DR HSP; P11362; 1FGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyfK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 545 AA; 60899 MW; 140DCB8CCA0F8AF4 CRC64;
```

```
Query Match 24.1%; Score 1645; DB 15; Length 545;
Best Local Similarity 54.9%; Pred. No. 5.5e-116;
Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15;

QY 578 GGEADQCVACAHYKDPFCVACRCPGVKPDLSYPIWKFPDEEGACQPCPINTHSCVDL 637
DB 1 GP-DHCWKCAHFIDGPHCVKACPAVLGENDTL-VWKYADANAVCQLCHPNCRTGCKGP 57
QY 638 DKKGCPAEQRASPLTSIVSAVV-GILLVVVLGVVFGILIKRQOKIRKYTMRLLOETEL 696
DB 58 GLEGCP---NGSKTPSIAAGVVGGLCLVVVGLGIGLYLRR-HIVRKETLRLLOEREL 113
QY 697 VEPLTPSGAMPNQAMRLKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVL 756
DB 114 VEPLTPSGEAPNQAHRLKETEFKKVKVLGFGAGFTVYKGLWIPEGEKVTIPVAIKEL 173
QY 757 ENTPSKANKEILDEAYVMAGVSPVSRLLGICLTSTVOLVTLQMPYGCCLDHVRENRL 816
DB 174 EATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVOLVTLQMPYGCCLDYREHKDN 233
QY 817 LGSQDLNWCQIAKGMYSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETE 876
DB 234 IGSQYLLNWCQIAKGMNLYLEERHVRDLAARNVLKTPQDVKITDFGLAQQLGADEKE 293
QY 877 YHADGKVPKMALESILRRFTHQSDVWSYGVTVVWELMTFGAKPYDGPAPREIPDLLE 936
DB 294 YHAGGKVPKMALESILHRYTHQSDVWSYGVTVVWELMTFGSKPYDGPASEISSVLE 353
QY 937 KGERLPQPPICITIDVYIMVKCMWIDSECRPRFRELSEFSRMARDPQRFVVIQ-NEDLG 995
DB 354 KGERLPQPPICITIDVYIMVKCMWGDASRPKPRFRELIAEFSKWARDPPRYLVIQGDERMH 413
QY 996 PASPLDSTFYRSLLEDDDMGLVDAEYLVPOQGFCDPAPGAGGMVHRHRSSTRSG 1055
DB 414 LPSPTDSKFYRTLMEEDMEDIVDAEYLVPHQGF-----NSPST---- 454
```

```

QY 1056 GGLTLGLPSEBAPRSL-----APSEGAGSDVFDGLGMAAGLQSLPHTDPSPLQ 1110
Db 455 -----SFTPLLSLSATSNSATNCIDRNG-----H----- 481
QY 1111 RYSEDPTVPLPSETDGVAPLTCSPQPEYVQPDVRPQPPSPREGPLPAARAGAT-LER 1169
Db 482 -----PVREDGFL-----PAPEYVQ--LMPKPESTAMVQIQIYVYISLTAISK 523
QY 1170 AKTLPSPKGVKVDVFAFGGAVENPEYL 1197
Db 524 LPMDSRYQN-----SHSTAVDNPEYL 544

RESULT 15
Q9WVF5 Q9WVF5 PRELIMINARY; PRT; 655 AA.
AC Q9WVF5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Epidermal growth factor receptor (Epidermal growth factor receptor
DE isoform 3).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RA Reiter J.L., Threadgill D.W., Danielson A.J., Schell C.,
RA Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
RA Mahle N.J.;
RT "Alternative Transcripts from the Human and Mouse EGFR Genes Encode
RT Carboxy-Terminal Truncated Receptors.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEV7AC;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Sincclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Mahle N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egfr transcripts encoding truncated receptor
RT isoforms.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sakai H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cdna collection.";
RL Nature 409:685-690(2001).
DR EMBL; AF124513; AAD44149.1; -.
DR EMBL; AF275366; AAG28047.1; -.

```

```

DR EMBL; AF275364; AAG28047.1; JOINED.
DR EMBL; AF275365; AAG28047.1; JOINED.
DR EMBL; AK004944; BAB23688.1; -.
DR EMBL; AK004883; BAB23641.1; -.
DR EMBL; AK004911; BAB23662.1; -.
DR MGD; MGI:95294; Egfr.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR SMART; SM00261; FU; 3.
KW Receptor.
SQ SEQUENCE 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;

Query Match 22.0%; Score 1500.5; DB 11; Length 655;
Best Local Similarity 43.2%; Pred. No. 6.4e-105;
Matches 281; Conservative 98; Mismatches 226; Indels 45; Gaps 8;

QY 11 LLLALLPPGAA-----STQVCTFNFTVFWLVPKVSASHLELYQGCVVQ 57
Db 14 LITALCAAGALEEKVKCGTSNRLTGLCTFEDHFLSL-----QRYNNCEVL 62
QY 58 GNLEUTYLTNASLSPFDIQEVQGVYLIAHNOVQVPLQRLIRVIRVGTOLFEDNYALAVL 117
Db 63 GNLEITYVQRNYDLSPFKTIQEVAGVYLIATNTVERIPLENLQIRGNALYENTYALAIL 122
QY 118 DNGDPLNNTPTVTGASPGGLRELSLTELKGGVLIQENPOLCYQDTILAKDI----F 173
Db 123 SN-----YGNRTGLRELPMNLQELIGAVRFSNNPILCNMDTIQWRDIVQNVF 172
QY 174 HKNQALALTLDITNRSRACHPCSPCKGSRGCGESSEDCQSLTRTVTCAGGCA-RCKGPLP 232
Db 173 MSNMSNDL-----QSHSSCPKCDPSPGSCWGGEENCQKLTIIICAQCSHRCGRSP 228
QY 233 TDCHEQCAAGCTGPKHSDCLACLHPNHSGLI CELHCPALVYNTVNTDTFESMPNDEGRYTFG 292
Db 229 SDCCNQCACAGCTGPRSDCLVCQKFDEATCKDTCPLMLYNTPTTYQMDVNPGEKYSFG 288
QY 293 ASCVTACPNYLSLTDVSGCTLVCPLNHNOVTAEDGTORCEKCKPCARVCYGLGMEHLRE 352
Db 289 ATCVKCKPRNVVYTDHSGCVRAGPDYEV-BEDGIRKCKCKDGPCKKVCNGIGIGEFKD 347
QY 353 VRAVTSANIQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQEQLOVFTLEETIGLYI 412
Db 348 TLSINATNIKHFYCTAISGLHILPVPFKGDSFTRTPLDPRELEILKTVKEITGFLLI 407
QY 413 SAWPDSLPLSVFQNLQVIRGRILHNGAYSLTQGLGISWGLRSLRSLRGLSLALIHNT 472
Db 408 QAMPDNWTDLHAFENLEIRGTQKHQGFSLAVVGLNITSLGLRSLKEISDGVIIISGNR 467
QY 473 HLCFVHTVPMDOLFRNPHQALLHTANRPEDECVGEGLACHOLCARGCHGPGPTQCVNCS 532
Db 468 NLCYANTINWKLFGTPNOKTKIMNRAEDCKAVNHVNCNPLCSSSGCGWGPEDRDCVSCQ 527
QY 533 QFLRGQECVEECVRLQGLPREYVNAHRLCPHPEQOPQNGSVTCFQPEADQCVACARYKD 592
Db 528 NVSRGECVEKCNILLEGEPREFENSECICQHPCLPQAMNITCTCRGPDNCICQAHYID 587
QY 593 PPFVCARCSGVKPDLSYMPIWKFPPDEGACQPCPINCNTCHSCVDLDDKGC 642
Db 588 GPHCVKTCIPAGIMGENNTL-VMKYADANNVCHLCHANCTYGCAGPGLOGC 636

```

Search completed: July 22, 2003, 09:00:01
Job time : 54.3575 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:11:38 ; Search time 37.9774 Seconds
(without alignments)
4403.399 Million cell updates/sec

Title: SEQ4-5-25-14
Perfect score: 6812
Sequence: 1 MELALCRWGLLLALLPPGA.....TFKGTPTAENPEYGLDVPV 1255

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6700	98.4	1255	21	AAV92620 Human heregulin 2
2	6700	98.4	1255	22	AAE12130 Human tyrosine kin
3	6700	98.4	1255	22	AAB60167 HER2 transgene pla
4	6700	98.4	1255	23	AAU74545 Human HER2 (erbB2)
5	6694	98.3	1255	17	AAW01111 HER-2/neu protein.
6	6694	98.3	1255	20	AAW92406 Human HER-2/neu on
7	6694	98.3	1255	21	AAW21198 Human HER-2/neu pr
8	6694	98.3	1255	21	AAV84780 Amino acid sequenc
9	6694	98.3	1255	22	AAB85458 Human HER-2/neu pr
10	6694	98.3	1255	22	AAG88267 HER2/neu amino aci

11	6694	98.3	1255	23	AAE24067 Human Her-2 protei
12	6694	98.3	1255	23	AAE20479 Human Her-2/neu pr
13	6694	98.3	1255	23	AAE51143 Human Her-2/neu on
14	6694	98.3	1255	23	AAU77114 Human Her-2/neu po
15	6651	97.6	1433	14	AAU39568 Sequence of c-erbB
16	6572	96.5	1223	23	AAU98923 Human breast cance
17	6377	93.6	1200	21	AAE21208 Human HER-2/neu pr
18	5886.5	86.4	1256	21	AAE21199 Rat Her-2/neu prot
19	5886.5	86.4	1256	23	AAE51144 Rat Her-2/neu onco
20	5861.5	86.0	1256	21	AAE21206 Mouse Her-2/neu pr
21	5861.5	86.0	1256	22	AAG62860 Amino acid sequenc
22	5861.5	86.0	1256	23	AAE51151 Mouse Her-2/neu on
23	4780	70.2	919	21	AAE21203 Human HER-2/neu fu
24	4780	70.2	919	23	AAE51148 Her-2/neu extracel
25	4011.5	58.9	920	23	AAE51152 Mouse Her-2/neu ex
26	4011.5	58.9	926	23	AAE51153 Mouse Her-2/neu ex
27	3664	53.8	712	21	AAE21204 Human HER-2/neu fu
28	3664	53.8	712	23	AAE51149 Her-2/neu extracel
29	3518	51.6	782	18	AAW19764 Her2-GM-CSF immuno
30	3516	51.6	553	21	AAE21200 Extracellular HER-
31	3516	51.6	553	23	AAE51145 Human ErbB2 oncopr
32	3478	51.1	545	22	AAE60408 Human ErbB2 extrac
33	3478	51.1	545	22	AAE61593 Human ErbB2 extrac
34	3413	50.1	551	21	AAE44993 DC9cFv-erbB2EC fu
35	3310	48.6	524	11	AAE08222 Extracellular port
36	3132	46.0	1210	21	AAE50616 Amino acid sequenc
37	3132	46.0	1210	21	AAE19259 Human EGF receptor
38	3132	46.0	1210	23	AAE23019 Human Her-1 protei
39	3132	46.0	1210	23	AAE50768 Human epidermal gr
40	3130	45.9	1210	22	AAE68420 Amino acid sequenc
41	3091	45.4	1210	23	AAE51768 Human epidermal gr
42	3084	45.3	583	23	AAE20483 Human protein for
43	3084	45.3	587	23	AAE20481 Human protein for
44	3083	45.3	589	23	AAE20484 Human protein for
45	3083	45.3	600	23	AAE20482 Human protein for

ALIGNMENTS

RESULT 1
AAV92620
ID AAV92620 standard; Protein; 1255 AA.
XX
AC AAV92620;
XX
DT 10-AUG-2000 (first entry)
XX
DE Human heregulin 2 (Her2).
XX
KW Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;
KW self-protein; cancer; breast cancer; prostate cancer;
KW cell-associated peptide antigen; foreign epitope.
XX
OS Homo sapiens.
XX
FH Key
FH Domain
FT
FT /label= N-terminal
FT /note= "mature polypeptide"
FT Region
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT Region
FT /label= insertion_region
FT /note= "suitable for foreign epitope insertion"
FT Region
FT /label= insertion_region
FT /note= "suitable for foreign epitope insertion"
FT Region
FT /label= insertion_region
FT /note= "suitable for foreign epitope insertion"
FT Domain

FT Region /label= Cysteine_rich_domain
 210..224
 /label= insertion region
 /note= "suitable for foreign epitope insertion"
 FT Region 250..264
 /label= insertion region
 /note= "suitable for foreign epitope insertion"
 FT Domain 324..483
 /label= Ligand_binding_domain
 325..339
 /label= insertion region
 /note= "suitable for foreign epitope insertion"
 FT Region 369..383
 /label= insertion region
 /note= "suitable for foreign epitope insertion"
 FT Region 465..479
 /label= insertion region
 /note= "suitable for foreign epitope insertion"
 FT Domain 484..623
 /label= Cysteine_rich_domain
 579..593
 /label= insertion region
 /note= "suitable for foreign epitope insertion"
 FT Domain 624..654
 /label= Transmembrane_domain
 632..652
 /label= insertion region
 /note= "suitable for foreign epitope insertion"
 FT Region 653..667
 /label= insertion region
 /note= "suitable for foreign epitope insertion"
 FT Domain 655..1010
 /label= Tyrosine_kinase_domain
 661..675
 /label= insertion region
 /note= "suitable for foreign epitope insertion"
 FT Region 695..709
 /label= insertion region
 /note= "suitable for foreign epitope insertion"
 FT Region 710..730
 /label= insertion region
 /note= "suitable for foreign epitope insertion"
 FT Domain 1011..1235
 /label= C-terminal_domain
 XX WO200020027-A2.
 XX 13-APR-2000.
 XX 05-OCT-1999; 99WO-DK00525.
 XX 05-OCT-1998; 98DK-0001261.
 XX 20-OCT-1998; 98US-0105011.
 XX (MEBI-) M & E BIOTECH AS.
 XX Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I;
 PI Gautam A, Birk P, Karlsson G;
 XX WPI; 2000-349917/30.
 DR N-PSDB; AAA09455.
 XX Inducing immune responses to weakly immunogenic, tumor associated
 PT peptide antigens for the treatment of breast and prostate cancer
 XX
 PS Claim 62; Page 193-198; 220pp; English.
 XX
 CC This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of
 CC Her2 can be used in the claimed method as an autovaccine to induce a CTL
 CC response. Subdominant CTL epitopes, antibody binding regions and
 CC cysteine residues involved in disulfide bonds are preserved in the
 CC immunogenized forms. Regions suitable for the insertion of foreign T
 CC helper epitopes were identified (see features table). The method

CC is used for inducing immune responses against weakly immunogenic
 CC cell-associated peptide antigens (PA) such as those associated with
 CC cancers (self-proteins), e.g. human prostate specific membrane antigen
 CC (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).
 CC The method comprises effecting simultaneous presentation by antigen
 CC producing cells (APCs) of the animals immune system of: (1) at least 1
 CC CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1
 CC B-cell group derived from the cell-associated PA; and (2) at least 1
 CC first T helper cell group which is foreign to the animal. Analogues of
 CC human PSM, human Her2 and human/murine FGF8b comprising a substantial
 CC part of all known and predicted CTL and B-cell epitopes of the respective
 CC PA and including at least one foreign T helper epitope are also claimed.
 CC The method is used to treat prostate, prostate/breast or breast cancer
 CC when the PA is human PSM, FGF8b and Her2, respectively.
 XX
 SQ Sequence 1255 AA;

Query Match 98.4%; Score 6700; DB 21; Length 1255;
 Best Local Similarity 98.4%; Pred. No. 0;
 Matches 1239; Conservative 4; Mismatches 8; Indels 8; Gaps 2;

QY 1 MELAALCRWGLLLALLPPGAASSTVCTFNNFTVSWFLRVKVSASHLE----LYQGCVV 56
 DB 1 MELAALCRWGLLLALLPPGAASSTVCTGDMK----LRLPASPETHLDLRLHYQSCV 56
 QY 57 QGNLELYLPTNASLSFLQDIOEVQGYVLIANQVROVPLQRLIRVGTQLPEDNYALV 116
 DB 57 QGNLELYLPTNASLSFLQDIOEVQGYVLIANQVROVPLQRLIRVGTQLPEDNYALV 116
 QY 117 LONGDPLNNTTPTVGASPGGLRELQRLSRLTEILKGVLIQRLNQLYQDILWKDIFHKN 176
 DB 117 LONGDPLNNTTPTVGASPGGLRELQRLSRLTEILKGVLIQRLNQLYQDILWKDIFHKN 176
 QY 177 NQALTLIDTNRSRACHPCSPCKSGRCWGESSEDCQSLTRTVCCAGCARCKGLPTDCC 236
 DB 177 NQALTLIDTNRSRACHPCSPCKSGRCWGESSEDCQSLTRTVCCAGCARCKGLPTDCC 236
 QY 237 HEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDFESNPNEGRTVFCASCV 296
 DB 237 HEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDFESNPNEGRTVFCASCV 296
 QY 297 TACPNYLTSDVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMBHLREVAV 356
 DB 297 TACPNYLTSDVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMBHLREVAV 356
 QY 357 TSANTQEFAGCKKIFGSLAFLPESFGDPSANTAPLOPEQLQVFETLEETGYLYISAWP 416
 DB 357 TSANTQEFAGCKKIFGSLAFLPESFGDPSANTAPLOPEQLQVFETLEETGYLYISAWP 416
 QY 417 DSLPDLVSFQNLQVIRGRILHNGAYSLTQGLGISWGLRLSRLSGSLALHNNHLCF 476
 DB 417 DSLPDLVSFQNLQVIRGRILHNGAYSLTQGLGISWGLRLSRLSGSLALHNNHLCF 476
 QY 477 VHTVPWDQLFRNPQALLHTANRPEDECYGEGLACHQLCARGHGWPGTQCVCNCSQFLR 536
 DB 477 VHTVPWDQLFRNPQALLHTANRPEDECYGEGLACHQLCARGHGWPGTQCVCNCSQFLR 536
 QY 537 GQECVBECHVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDPDFC 596
 DB 537 GQECVBECHVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDPDFC 596
 QY 597 VARCPGVKPDLSYMPKFPDEBACQPCPINCTHSCVDLDDKGPAPQASPLTSIYS 656
 DB 597 VARCPGVKPDLSYMPKFPDEBACQPCPINCTHSCVDLDDKGPAPQASPLTSIYS 656
 QY 657 AVVGILLVVVLGVWFVILIKRQKIRKYTMRLRLQETELVEPLTPSGAMPNQAMRILK 716
 DB 657 AVVGILLVVVLGVWFVILIKRQKIRKYTMRLRLQETELVEPLTPSGAMPNQAMRILK 716
 QY 717 ETELKRVKVLGSGAGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAG 776
 DB 717 ETELKRVKVLGSGAGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAG 776


```

QY 717 ETELKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAG 776
DB 717 ETELKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAG 776
QY 777 VGSPPYVRLIGICLTSTVQLVTQMPYGCCLLDHVRENRGLSGQDLLNWCQIAGKMSYL 836
DB 777 VGSPPYVRLIGICLTSTVQLVTQMPYGCCLLDHVRENRGLSGQDLLNWCQIAGKMSYL 836
QY 837 EDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKKNWALSILR 896
DB 837 EDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKKNWALSILR 896
QY 897 RRFTHQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGBRLPQPPTCTIDVYMIW 956
DB 897 RRFTHQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGBRLPQPPTCTIDVYMIW 956
QY 957 KCMWIDSECHPRFRELVSERFARDPQFVVIQNEIDLGPASPLDSTFYRSLLEDMDMGD 1016
DB 957 KCMWIDSECHPRFRELVSERFARDPQFVVIQNEIDLGPASPLDSTFYRSLLEDMDMGD 1016
QY 1017 LVDAEYLVPOQGFCCPDPAAGAGVHHRHRSSTRSGGDLTLGLEPSEERAPRPLA 1076
DB 1017 LVDAEYLVPOQGFCCPDPAAGAGVHHRHRSSTRSGGDLTLGLEPSEERAPRPLA 1076
QY 1077 PSEGAGSDVFDGLGMAAKGLQSLPTHDPSPLOQYSEDPTVLPSETDGYVAPLTCSPQ 1136
DB 1077 PSEGAGSDVFDGLGMAAKGLQSLPTHDPSPLOQYSEDPTVLPSETDGYVAPLTCSPQ 1136
QY 1137 PEYVNPQDVRPQPPSPREGPLPAARAGATLERAKTILSPGKNGVVDVFAFGAVENPEY 1196
DB 1137 PEYVNPQDVRPQPPSPREGPLPAARAGATLERAKTILSPGKNGVVDVFAFGAVENPEY 1196
QY 1197 LTPQGGAAPOHPHPPAFSPAFDNLXYWDQDPPPERGAPPSTFKTPTAENPEYLGLDVVP 1255
DB 1197 LTPQGGAAPOHPHPPAFSPAFDNLXYWDQDPPPERGAPPSTFKTPTAENPEYLGLDVVP 1255

RESULT 3
AAB60167
ID AAB60167 standard; Protein; 1255 AA.
XX AC AAB60167;
XX DT 03-APR-2001 (first entry)
XX DE HER2 transgene plasmid construct encoded protein.
XX KW Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
XX OS Homo sapiens.
XX OS Synthetic.
XX PN WO200100244-A2.
XX PD 04-JAN-2001.
XX PF 23-JUN-2000; 2000WO-US17229.
XX PR 25-JUN-1999; 99US-0141316.
XX PR 16-MAR-2000; 2000US-0189844.
XX PA (GETH ) GENENTECH INC.
XX PI Erickson S, Schwall R;
XX PR N-PSDB; AAF24297.
XX PT Treating tumors, particularly breast cancers, which overexpress an ErbB
XX PT receptor and does not respond to an anti-ErbB antibody, comprises
XX PT conjugating the antibody to a maytansinoid -
XX

```

Example 3; Fig 4; 92pp; English.

The present invention provides a method of treating cancer by administering a conjugate of anti-ErbB antibody with a maytansinoid. In particular, the antibody is directed against ErbB2 (also known as HER2 and p185neu). The method is particularly useful in the treatment of breast, ovarian, stomach, endometrial, salivary gland, lung, kidney, colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.

Sequence 1255 AA;

Query Match 98.4%; Score 6700; DB 22; Length 1255;
 Best Local Similarity 98.4%; Pred. No. 0;
 Matches 1239; Conservative 4; Mismatches 8; Indels 8; Gaps 2;

```

QY 1 MELAALCRWGLLALLPPGAASVTCTFNNFTVSWLRVPKVSASHLE---LYOGCQV 56
DB 1 MELAALCRWGLLALLPPGAASVTCTGDMK----LRLPASPTHLDMLRHLYGQCV 56
QY 57 QGNLELTYPNTNASLSFLQDIOEVQYVLIANQVRQVPLQRLRIVRGTLQFEDNYALAV 116
DB 57 QGNLELTYPNTNASLSFLQDIOEVQYVLIANQVRQVPLQRLRIVRGTLQFEDNYALAV 116
QY 117 LDNGDPLNNTTPTVTCASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHK 176
DB 117 LDNGDPLNNTTPTVTCASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHK 176
QY 177 NQALATLDTNRSRACHPCSPCKGRCWGESSEDCQSLTRTVCAAGCARCKPLPTDCC 236
DB 177 NQALATLDTNRSRACHPCSPCKGRCWGESSEDCQSLTRTVCAAGCARCKPLPTDCC 236
QY 237 HQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGYTFGASCV 296
DB 237 HQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGYTFGASCV 296
QY 297 TACPNNYLTVDGSCCTLVCPLNQVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRAV 356
DB 297 TACPNNYLTVDGSCCTLVCPLNQVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRAV 356
QY 357 TSANIQEAFAGCKKIFGSLAFIPESFDGDPASNTAPLOPELOVFTLEITGYLISAWP 416
DB 357 TSANIQEAFAGCKKIFGSLAFIPESFDGDPASNTAPLOPELOVFTLEITGYLISAWP 416
QY 417 DSLPDLVSFQNLQVIRGRILHNGAYSILTQGLGISWLGRLSRLGSLGLALHNNTHLCF 476
DB 417 DSLPDLVSFQNLQVIRGRILHNGAYSILTQGLGISWLGRLSRLGSLGLALHNNTHLCF 476
QY 477 VHTVPWDQLFRNPHQALLHTANRPEDECVGEGLAGHQLCARGHGWPGPTQCVNCSQFLR 536
DB 477 VHTVPWDQLFRNPHQALLHTANRPEDECVGEGLAGHQLCARGHGWPGPTQCVNCSQFLR 536
QY 537 GOECVEECRVLQGLPREYNARHCLPCHPECOPQNGSVTCFGEADQCVACAHYKDPFPC 596
DB 537 GOECVEECRVLQGLPREYNARHCLPCHPECOPQNGSVTCFGEADQCVACAHYKDPFPC 596
QY 597 VARCPGVKPDLSYMPIWKFPDEEGACOPCPINCTHSCVDLDDKGCAPAEORASPLTSIVS 656
DB 597 VARCPGVKPDLSYMPIWKFPDEEGACOPCPINCTHSCVDLDDKGCAPAEORASPLTSIVS 656
QY 657 AVVGILLVVVLGVVFGILIKRRQKIRKYTWERLLQETELVEPLTPSGAMPNQAQMRILK 716
DB 657 AVVGILLVVVLGVVFGILIKRRQKIRKYTWERLLQETELVEPLTPSGAMPNQAQMRILK 716
QY 717 ETELKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAG 776
DB 717 ETELKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAG 776
QY 777 VGSPPYVRLIGICLTSTVQLVTQMPYGCCLLDHVRENRGLSGQDLLNWCQIAGKMSYL 836
DB 777 VGSPPYVRLIGICLTSTVQLVTQMPYGCCLLDHVRENRGLSGQDLLNWCQIAGKMSYL 836
QY 837 EDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKKNWALSILR 896
DB 837 EDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKKNWALSILR 896

```

Db 837 EDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIKWMALESILR 896
QY 897 RRFTHQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGBRLPQPPICTIDVYMIW 956
Db 897 RRFTHQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGBRLPQPPICTIDVYMIW 956
QY 957 KCMWIDSECHPRELVSFESRWARDPQRFVVIQNEEDLGPASPLDSTFYRSLLDDMGD 1016
Db 957 KCMWIDSECHPRELVSFESRWARDPQRFVVIQNEEDLGPASPLDSTFYRSLLDDMGD 1016
QY 1017 LVDAAEYLVPOQGFPCDPAPGAGGMVHRHRSSTRSGGDLTLGLEPSEEAAPSPLA 1076
Db 1017 LVDAAEYLVPOQGFPCDPAPGAGGMVHRHRSSTRSGGDLTLGLEPSEEAAPSPLA 1076
QY 1077 PSEGAGSDVFDGLMGAAKGLQSLPHTDPSLQRYSEDTVPLPSETDGYVAPLTCSPQ 1136
Db 1077 PSEGAGSDVFDGLMGAAKGLQSLPHTDPSLQRYSEDTVPLPSETDGYVAPLTCSPQ 1136
QY 1137 PEYVNPQDVRPQPPSPREGPLPAARAGATLEAKTILSPKNGVVKDVFAGGAVENPEY 1196
Db 1137 PEYVNPQDVRPQPPSPREGPLPAARAGATLEAKTILSPKNGVVKDVFAGGAVENPEY 1196
QY 1197 LTPQGAAPQPHPPPAFSPAFDNLXYWDQDPPPERGAPPSTFKTPTAENPEYLGLDVPV 1255
Db 1197 LTPQGAAPQPHPPPAFSPAFDNLXYWDQDPPPERGAPPSTFKTPTAENPEYLGLDVPV 1255
RESULT 4
AAU74545
ID AAU74545 standard; Protein; 1255 AA.
XX
AC AAU74545;
XX
DT 23-APR-2002 (first entry)
XX
DE Human HER2 (ErbB2) polypeptide.
XX
KW Human; HER2; ErbB; epidermal growth factor receptor;
KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
KW glial disorder; astrocytal disorder; hypothalamic disorder;
KW glandular disorder; macrophagal disorder; epithelial disorder;
KW stromal disorder; blastocoealic disorder; inflammatory disorder;
KW angiogenic disorder; immunological disorder.
XX
OS Homo sapiens.
XX
PN US2002001587-A1.
XX
PD 03-JAN-2002.
XX
PF 16-MAR-2001; 2001US-0811123.
XX
PR 16-MAR-2000; 2000US-189844P.
PR 05-OCT-2000; 2000US-238327P.
XX
PA (ERIC/) ERICKSON S.
PA (SCHW/) SCHWALL R.
PA (SLIW/) SLIWKOWSKI M.
XX
PI Erickson S, Schwall R, Sliwowski M;
XX
DR WPI; 2002-163686/21.
DR N-PSDB; ABK14058.
XX
PT Treating tumour characterised by overexpression of epidermal growth
PT factor receptor, ErbB or cancer in mammal, comprises administering
PT anti-ErbB antibody-maytansinoid conjugate to the mammal -
XX
PS Example 3; Fig 7; 93pp; English.
XX
CC The invention relates to treating a tumour in a mammal, where the tumour

CC is characterised by the overexpression of an epidermal growth factor
CC receptor (ErbB) and does not respond or responds poorly, to treatment
CC with an anti-ErbB antibody, comprising administering to the mammal an
CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for
CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas, a
CC prostate and bladder, preferably breast cancer. The breast cancer is a
CC metastatic breast cancer or an aggressive form of metastatic breast
CC cancer which overexpresses ErbB2. The method is also useful for treating
CC neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal,
CC epithelial, stromal, blastocoealic, inflammatory, angiogenic and
CC immunological disorders. This sequence represents the human HER2 (ErbB2)
CC polypeptide of the invention.
XX
SQ Sequence 1255 AA;
Query Match 98.4%; Score 6700; DB 23; Length 1255;
Best Local Similarity 98.4%; Pred. No. 0;
Matches 1239; Conservative 4; Mismatches 8; Indels 8; Gaps 2;
QY 1 MELAALCRWGLLALLPPGAASTQVCTFNNFTVSFWLRVPKVSASHLE----LYQSCQV 56
Db 1 MELAALCRWGLLALLPPGAASTQVCTGDMK---LRLPASPEETHLDMRLHLYQSCQV 56
QY 57 QCNLELTYPNNAISFLDIOEVQGYVLIANQVRQVPLQRLIRVRGTQLFEDNVALAV 116
Db 57 QCNLELTYPNNAISFLDIOEVQGYVLIANQVRQVPLQRLIRVRGTQLFEDNVALAV 116
QY 117 LDNGDPLANNTPVTGASPGGLRELOLRSITELTKGGVLIQRPOLCYQDTILWKDIFHKN 176
Db 117 LDNGDPLANNTPVTGASPGGURELQRLSITELTKGGVLIQRPOLCYQDTILWKDIFHKN 176
QY 177 NQLALTLIDTIRSRACHPCSPMKGSRGWGESSEDCQSLTRTVACGGACRCKGPIPTDCC 236
Db 177 NQLALTLIDTIRSRACHPCSPMKGSRGWGESSEDCQSLTRTVACGGACRCKGPIPTDCC 236
QY 237 HQCAAGCTGPKHSDCLACLHPNHSGICELHCPALVTYNTDTFESMPNPEGRYTGASCV 296
Db 237 HQCAAGCTGPKHSDCLACLHPNHSGICELHCPALVTYNTDTFESMPNPEGRYTGASCV 296
QY 297 TACPYNILSTDVGSCTLCPLHNOVTAEDGTORCEKSKPCARVCYGLGMEHLREVAV 356
Db 297 TACPYNILSTDVGSCTLCPLHNOVTAEDGTORCEKSKPCARVCYGLGMEHLREVAV 356
QY 357 TSANIQEFAGCKKIFGSLAFLPESPDGPASNTAPLOPELOVVFETLEITGYLISAMP 416
Db 357 TSANIQEFAGCKKIFGSLAFLPESPDGPASNTAPLOPELOVVFETLEITGYLISAMP 416
QY 417 DSLPDLVSFQNLQVIRGRILHNGAYSILTQGLGISWLGRLSRLRELSGLALIHNNTHLCF 476
Db 417 DSLPDLVSFQNLQVIRGRILHNGAYSILTQGLGISWLGRLSRLRELSGLALIHNNTHLCF 476
QY 477 VHTVPMDQLFRNPQALLHTANRPEDECVGEGELACHOLCARGHCWGPGTQCNCVSQFLR 536
Db 477 VHTVPMDQLFRNPQALLHTANRPEDECVGEGELACHOLCARGHCWGPGTQCNCVSQFLR 536
QY 537 GOECVBECKRVLOGLPREVYNARHCLPCHPECOPOGNSVTCFPGPADQCVACAHYKDPFPC 596
Db 537 GOECVBECKRVLOGLPREVYNARHCLPCHPECOPOGNSVTCFPGPADQCVACAHYKDPFPC 596
QY 597 VARCPGSKVPDLSYMPIWKFPDEGACQPCPINCTHSCVLDLDDKGPQASPLTSIVS 656
Db 597 VARCPGSKVPDLSYMPIWKFPDEGACQPCPINCTHSCVLDLDDKGPQASPLTSIVS 656
QY 657 AVVIGILLVVVLGVWFGILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRILK 716
Db 657 AVVIGILLVVVLGVWFGILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRILK 716
QY 717 ETELAKVKVLSSGAFGTGVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMAG 776
Db 717 ETELAKVKVLSSGAFGTGVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMAG 776
QY 777 VGSYPVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRRGLSGQDLNWCQIAKMSYL 836

Db 777 VGSPPVSRLLGICLTSTVQLVTLQMPYGCCLLDHVRNRRGLSGQDLNLCMQIAKMSYL 836
QY 837 EDVRLVHRDLAARNVLKSNHVKITDGLARLLDIDETEHADGGKVPKKNWALSILR 896
Db 837 EDVRLVHRDLAARNVLKSNHVKITDGLARLLDIDETEHADGGKVPKKNWALSILR 896
QY 897 RRETHQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGRLLPQPICTIDVYMIW 956
Db 897 RRETHQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGRLLPQPICTIDVYMIW 956
QY 957 KCWMDSECRPRELVSEFSRWARDPQRFVVIQNEEDLGPASPLDSTFYRSLLDDMDGD 1016
Db 957 KCWMDSECRPRELVSEFSRWARDPQRFVVIQNEEDLGPASPLDSTFYRSLLDDMDGD 1016
QY 1017 LVDAEYLVPOQFFCPDPAAGGVHHRSSSTRSGGDLTLGLESEEAAPRSLA 1076
Db 1017 LVDAEYLVPOQFFCPDPAAGGVHHRSSSTRSGGDLTLGLESEEAAPRSLA 1076
QY 1077 PSEGASDVFDGLGMAAKGLQSLPHTDPSPLQRYSEDTVPLPSETDGYVAPLTCSPQ 1136
Db 1077 PSEGASDVFDGLGMAAKGLQSLPHTDPSPLQRYSEDTVPLPSETDGYVAPLTCSPQ 1136
QY 1137 PEYVNPQDVRPQPPREGPLPAARPAAGATLERAKTSLPGKNGVVKDVFAFGAVENPEY 1196
Db 1137 PEYVNPQDVRPQPPREGPLPAARPAAGATLERAKTSLPGKNGVVKDVFAFGAVENPEY 1196
QY 1197 LTPQGAAPQPHPPAFSPAFDNLVYWDQPPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255
Db 1197 LTPQGAAPQPHPPAFSPAFDNLVYWDQPPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 5

AAW01111 standard; Protein; 1255 AA.

AC AAW01111;

DT 01-JAN-1997 (first entry)

XX HER-2/neu protein.

XX HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
KW breast cancer; ovary cancer; colon cancer; lung cancer;
KW prostate cancer; immunisation; tumour; vaccine; vector.

XX Homo sapiens.

XX Key Location/Qualifiers
FH Domain 676..1255
FT /label= Intracellular domain
FT /note= "claimed domain, useful for immunisation"

XX W09630514-A1.
XX 03-OCT-1996.
XX 28-MAR-1996; 96WO-0501689.
XX 31-MAR-1995; 95US-0414417.
XX (UNIW) UNIV WASHINGTON.
XX Cheever MA, Disis ML;
XX WPI; 1996-455361/45.
XX N-PSDB; AAT40739.

XX DNA encoding HER-2-neu poly.peptide(s) - used for prevention or
PT treatment of malignancies with which the HER-2/neu oncogene is
PT associated
XX Claim 2; Page 56-61; 71pp; English.

XX Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is
CC the product of the HER-2/neu oncogene (see also AAT40739). The
CC protein is over-expressed in various cancers, including breast,
CC ovarian, colon, lung and prostate. The intracellular domain of the
CC protein can be used to immunise an animal against a malignancy with
CC which the oncogene is associated. The polypeptide can be produced
CC in transformed host cells for use in immunisation. Alternatively,
CC animal cells are transfected in vivo or ex vivo with a viral vector
CC that directs expression of the polypeptide.

XX SQ Sequence 1255 AA;

Query Match 98.3%; Score 6694; DB 17; Length 1255;
Best Local Similarity 98.3%; Pred. No. 0;

Matches 1237; Conservative 5; Mismatches 9; Indels 8; Gaps 2;

QY 1 MELAALCRWGLLLALLPPGAASTVCTFNNFTVSFWMRVKVSASHLE----LYQSCQVW 56
Db 1 MELAALCRWGLLLALLPPGAASTVCTGTDK---LRLPASPEHLDMLRLHYQGCQVW 56

QY 57 QGNLELTYPNNTVPTGASPGGLRELOLRSLTEILKGVLIQRPOLCYQDTILWKDIFHKN 176
Db 57 QGNLELTYPNNTVPTGASPGGLRELOLRSLTEILKGVLIQRPOLCYQDTILWKDIFHKN 176

QY 117 LQNGDPLNNTVPTGASPGGLRELOLRSLTEILKGVLIQRPOLCYQDTILWKDIFHKN 176
Db 117 LQNGDPLNNTVPTGASPGGLRELOLRSLTEILKGVLIQRPOLCYQDTILWKDIFHKN 176

QY 177 NQALTLTIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRVCAGGCARCKGLPTDCC 236
Db 177 NQALTLTIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRVCAGGCARCKGLPTDCC 236

QY 237 HEQCAAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCV 296
Db 237 HEQCAAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCV 296

QY 297 TACPNYLVSTDVGSCTLVCPHNVQVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRV 356
Db 297 TACPNYLVSTDVGSCTLVCPHNVQVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRV 356

QY 357 TSANIQEFAGCKKIFGSLAFPESEFDGPASNTAPLQPEQLQVFETLEITGYLYISAMP 416
Db 357 TSANIQEFAGCKKIFGSLAFPESEFDGPASNTAPLQPEQLQVFETLEITGYLYISAMP 416

QY 417 DSLPDLVFNQNLQVIRGRILHNGAYSILTQGLGISWLGRLSRLSGSLALIHNTLHLCF 476
Db 417 DSLPDLVFNQNLQVIRGRILHNGAYSILTQGLGISWLGRLSRLSGSLALIHNTLHLCF 476

QY 477 VHTVPWDQLFRNPHQALLHTANRPEDECVGEGACHOLCARGHCWPGPTQCVCNCSQIFUR 536
Db 477 VHTVPWDQLFRNPHQALLHTANRPEDECVGEGACHOLCARGHCWPGPTQCVCNCSQIFUR 536

QY 537 GOECVEECRVLOGLPREYVYNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDPDPPFC 596
Db 537 GOECVEECRVLOGLPREYVYNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDPDPPFC 596

QY 597 VARCPGKVPDLVSYMPIWKFDPDEGACQPCPINCTHSCVDLDDKGCAPABQASPLTSIVS 656
Db 597 VARCPGKVPDLVSYMPIWKFDPDEGACQPCPINCTHSCVDLDDKGCAPABQASPLTSIVS 656

QY 657 AVVGILLVVLGVVFGILIKRQOKIRKYTWERLLQETELVEPLTPSGAMPNQAMRIUK 716
Db 657 AVVGILLVVLGVVFGILIKRQOKIRKYTWERLLQETELVEPLTPSGAMPNQAMRIUK 716

QY 717 ETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAG 776
Db 717 ETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAG 776

QY 777 VGSPPVSRLLGICLTSTVQLVTLQMPYGCCLLDHVRNRRGLSGQDLNLCMQIAKMSYL 836
Db 777 VGSPPVSRLLGICLTSTVQLVTLQMPYGCCLLDHVRNRRGLSGQDLNLCMQIAKMSYL 836

QY 837 EDVRLVHRDLAARNLVKSPNHVKITDPGLARLLDIDETEHADGGKVPKIMWALESIILR 896
 DB 837 EDVRLVHRDLAARNLVKSPNHVKITDPGLARLLDIDETEHADGGKVPKIMWALESIILR 896
 QY 897 RRTTHOSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGRLLPQPPICTIDVYMIW 956
 DB 897 RRTTHOSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGRLLPQPPICTIDVYMIW 956
 QY 957 KCWIMIDSECRPRFRELVSFSESRWARDPQRFVVIQNEEDLGPASPLDSTFYRSLLLEDDMGD 1016
 DB 957 KCWIMIDSECRPRFRELVSFSESRWARDPQRFVVIQNEEDLGPASPLDSTFYRSLLLEDDMGD 1016
 QY 1017 LVDAEYLYPQOQFFCPDPAAGAGGVHHRHRSSTRSGGDLTLGLEPSEEAAPRSPLA 1076
 DB 1017 LVDAEYLYPQOQFFCPDPAAGAGGVHHRHRSSTRSGGDLTLGLEPSEEAAPRSPLA 1076
 QY 1077 PSEGAGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDTVPPLPSETDGYVAPLTCSPQ 1136
 DB 1077 PSEGAGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDTVPPLPSETDGYVAPLTCSPQ 1136
 QY 1137 PEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTLPQKNGVVKDVFAFGGAVENPEY 1196
 DB 1137 PEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTLPQKNGVVKDVFAFGGAVENPEY 1196
 QY 1197 LTPQGGAAPOPHPPAPAFSADFNLVYWDQDPPPERGAPPSTFKTPTAENPEYLGLDVPV 1255
 DB 1197 LTPQGGAAPOPHPPAPAFSADFNLVYWDQDPPPERGAPPSTFKTPTAENPEYLGLDVPV 1255

RESULT 6

AAW92406

ID AAW92406 standard; Protein; 1255 AA.

XX

AC AAW92406;

XX

XX 21-APR-1999 (first entry)

XX

DE Human HER-2/neu oncogene protein.

XX

XX HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;

KW malignancy; treatment; tumour.

XX

XX Homo sapiens.

XX

FH Key Location/Qualifiers

FT Region 676..1255

FT /note= "region which elicits immune response"

XX

XX US5869445-A.

XX

XX 09-FEB-1999.

XX

XX 01-APR-1996; 96US-0625101.

XX

XX 01-APR-1996; 96US-0625101.

PR 17-MAR-1993; 93US-0033644.

PR 12-AUG-1993; 93US-0106112.

PR 31-MAR-1995; 95US-041417.

XX

XX (UNIW) UNIV WASHINGTON.

XX

XX Cheever MA, Disis ML;

XX

XX WPI; 1999-152835/13.

XX N-PSDB; AAX01912.

XX

XX Use of HER-2/neu polypeptides - for eliciting an immune response to
 PT an HER-2/neu associated malignancy, particularly for treating or
 PT preventing tumours

XX

PS Claim 3; Column 31-38; 26pp; English.

XX

XX This sequence represents the human HER-2/neu oncogene protein. A fragment

XX

CC of this protein is used in a method for eliciting or enhancing an immune
 CC response to HER-2/neu protein. The polypeptide can stimulate T cells and
 CC B cells to produce an immune response to the HER-2/neu protein. The
 CC method can be used for immunisation against a malignancy in which the
 CC HER-2/neu oncogene is associated and in the treatment of an existing
 CC tumour, or to prevent tumour occurrence or reoccurrence.

XX SQ Sequence 1255 AA;

Query Match 98.3%; Score 6694; DB 20; Length 1255;

Best Local Similarity 98.3%; Pred. No. 0;

Matches 1237; Conservative 5; Mismatches 9; Indels 8; Gaps 2;

QY 1 MELAALCRWGLLLALLPPGAASTVCTFNNFVSFVLRVPKVSASHLE----LYQGCQVV 56

DB 1 MELAALCRWGLLLALLPPGAASTVCTGTDK---LRLPASPEHLDMRLHLYQGCQVV 56

QY 57 QGNLELTYPNTASISFLQDIOEVQYVLIANQVQVPLQRLRIVRGTLFEDNYALAV 116

DB 57 QGNLELTYPNTASISFLQDIOEVQYVLIANQVQVPLQRLRIVRGTLFEDNYALAV 116

QY 117 LONGDPLNNTTPTVTCASPGGLRELQRLSLEILKGVLIQRPOLCYQDTILWKDIFHN 176

DB 117 LONGDPLNNTTPTVTCASPGGLRELQRLSLEILKGVLIQRPOLCYQDTILWKDIFHN 176

QY 177 NOLALTLDITNRSRACHPCSMKGRWGSEDCQSLTRTVCAAGCARCKGLPTDCC 236

DB 177 NOLALTLDITNRSRACHPCSMKGRWGSEDCQSLTRTVCAAGCARCKGLPTDCC 236

QY 237 HEQCAAGCTGPKHSDCLACLFHNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASC 296

DB 237 HEQCAAGCTGPKHSDCLACLFHNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASC 296

QY 297 TACPYNLYSTVGSCITVCPHLNQVETADGTQRCCKSKPCARVCYGLGMEHLREVAV 356

DB 297 TACPYNLYSTVGSCITVCPHLNQVETADGTQRCCKSKPCARVCYGLGMEHLREVAV 356

QY 357 TSANIQEFAGCKKIFGSLAFPESEFDGPASNTAPLQPEQLQVVFETLEITGYLYISAMP 416

DB 357 TSANIQEFAGCKKIFGSLAFPESEFDGPASNTAPLQPEQLQVVFETLEITGYLYISAMP 416

QY 417 DSLPDLSPVQNLQVIRGRILHNGAYSILQGLIGISWGLRLSRLGSLALIHNNHLFCF 476

DB 417 DSLPDLSPVQNLQVIRGRILHNGAYSILQGLIGISWGLRLSRLGSLALIHNNHLFCF 476

QY 477 VHTVWDQLFRNPHQALLHTANRDECEVGEGLACHOLCARGHCWGPPTQCVNCSQFLR 536

DB 477 VHTVWDQLFRNPHQALLHTANRDECEVGEGLACHOLCARGHCWGPPTQCVNCSQFLR 536

QY 537 GQECVBECEVRLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPC 596

DB 537 GQECVBECEVRLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPC 596

QY 597 VARPSPGVKPDLSYMPIWKFPDEGACQPCPINCTHSCVDLDDKCGCPABORASPLTSIYS 656

DB 597 VARPSPGVKPDLSYMPIWKFPDEGACQPCPINCTHSCVDLDDKCGCPABORASPLTSIYS 656

QY 657 AVVGILLVVVLGVVFEGLIKRQOKIRKYTWRLLOETELVEPLTPSGAMPNQAQMRILK 716

DB 657 AVVGILLVVVLGVVFEGLIKRQOKIRKYTWRLLOETELVEPLTPSGAMPNQAQMRILK 716

QY 717 ETELKRVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMAG 776

DB 717 ETELKRVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMAG 776

QY 777 VGSPVSRLLGLICTSTVQLVTQLMPYGLCLDHHVRENRGLSGDOLLNWCMIAGMSYL 836

DB 777 VGSPVSRLLGLICTSTVQLVTQLMPYGLCLDHHVRENRGLSGDOLLNWCMIAGMSYL 836

QY 837 EDVRLVHRDLAARNLVKSPNHVKITDPGLARLLDIDETEHADGGKVPKIMWALESIILR 896

DB 837 EDVRLVHRDLAARNLVKSPNHVKITDPGLARLLDIDETEHADGGKVPKIMWALESIILR 896

QY 897 RRTHQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPICTIDVYMW 956
DB 897 RRTHQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPICTIDVYMW 956
QY 957 KCMWIDSECHPRRELVSERMDARDQRFVWIONEDLGPASLDSTFYRSLLDDMDGD 1016
DB 957 KCMWIDSECHPRRELVSERMDARDQRFVWIONEDLGPASLDSTFYRSLLDDMDGD 1016
QY 1017 LVDAEYLVPOQGFCDPAPGAGGVHHRHRSSTRSGGDLTTLGLEPSEEAAPRSPLA 1076
DB 1017 LVDAEYLVPOQGFCDPAPGAGGVHHRHRSSTRSGGDLTTLGLEPSEEAAPRSPLA 1076
QY 1077 PSEGAGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDTVPLPSETDGYVAPLTCSPQ 1136
DB 1077 PSEGAGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDTVPLPSETDGYVAPLTCSPQ 1136
QY 1137 PEYVQPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVWVDVFAFGAVENPEY 1196
DB 1137 PEYVQPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVWVDVFAFGAVENPEY 1196
QY 1197 LTFQGGAAPOHPHPPAFSPAFDNLVYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
DB 1197 LTFQGGAAPOHPHPPAFSPAFDNLVYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255

RESULT 7

AAB21198
ID AAB21198 standard; protein; 1255 AA.

AC AAB21198;

XX 12-JAN-2001 (first entry)

DT Human HER-2/neu protein.

DE Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;

XX breast cancer; prostate cancer; ovarian cancer; lung cancer;

XX colon cancer.

OS Homo sapiens.

XX W0200044899-A1.

PN 03-AUG-2000.

XX 28-JAN-2000; 2000WO-US02164.

XX 29-JAN-1999; 99US-0117976.

XX (CORI-) CORIXA CORP.

PA (SMIK-) SMITHKLINE BEECHAM.

XX Cheever MA, Gheysen D;

PI WPI; 2000-505976/45.

XX N-PSDB; AAB89736.

XX HER-2/neu extracellular domain/phosphorylation domain fusion proteins

XX useful for vaccinating against breast, ovarian, colon, lung and

XX prostate cancers -

XX Claim 52; Fig 7; 128pp; English.

XX The present sequence is the human HER-2/neu protein. It is a member of

XX the tyrosine kinase family of receptor-like glycoproteins and shows

XX CC homology to the epidermal growth factor receptor (EGFR). It probably

XX CC plays a part in cell growth and/or differentiation. The HER-2/neu

XX CC gene is an oncogene. An HER-2/neu fusion protein comprising a

XX CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation

XX CC domain may be used to treat or prevent cancer by eliciting or

XX CC enhancing an immune response to the HER-2/neu protein. It may be used

XX CC to treat malignancies such as breast, ovarian, colon, lung and

XX CC prostate cancers, and may be used as an antigen to vaccinate against

CC these neoplasias.
XX
SQ Sequence 1255 AA;
Query Match 98.3%; Score 6694; DB 21; Length 1255;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1237; Conservative 5; Mismatches 9; Indels 8; Gaps 2;
QY 1 MELAALCRWGLLALLPPGAASCTVCTFNFTSVFWLRVPKVSASHLE---LYQCQCVV 56
DB 1 MELAALCRWGLLALLPPGAASCTVCTGTDMK----LRUPASPETHDMLRHLYYQCQCVV 56
QY 57 QGNLEITYLPTNASLSFLQDIOEVQGYVLIHAHQVRQVPLQRLRIVRGTLQFEDNYALAV 116
DB 57 QGNLEITYLPTNASLSFLQDIOEVQGYVLIHAHQVRQVPLQRLRIVRGTLQFEDNYALAV 116
QY 117 LONGDPLNNTTPTVGTASPGGLRELQRLSRLTEILKGVLIQRNPOLCYQDTILWKDIFHKN 176
DB 117 LONGDPLNNTTPTVGTASPGGLRELQRLSRLTEILKGVLIQRNPOLCYQDTILWKDIFHKN 176
QY 177 NOLALTLIDTNSRACHPCSPCKSGRCWGESSEDCQSLTRTVCAAGCARCKGLPTDCC 236
DB 177 NOLALTLIDTNSRACHPCSPCKSGRCWGESSEDCQSLTRTVCAAGCARCKGLPTDCC 236
QY 237 HEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCV 296
DB 237 HEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCV 296
QY 297 TACPNYILSTDVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRAV 356
DB 297 TACPNYILSTDVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRAV 356
QY 357 TSANIOEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEBITGYLISAMP 416
DB 357 TSANIOEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEBITGYLISAMP 416
QY 417 DSLPDLVSFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLRELGSGLALHNNHLCP 476
DB 417 DSLPDLVSFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLRELGSGLALHNNHLCP 476
QY 477 VHTVPWDQLFRNPHQALLHTANRPDECEVGEGLACHQLCARGHCWGPGTQCNCVQFLR 536
DB 477 VHTVPWDQLFRNPHQALLHTANRPDECEVGEGLACHQLCARGHCWGPGTQCNCVQFLR 536
QY 537 GOECVEECRVLOGLPREYVNAHCLPCHPECOQNGSVTCFGEADQCACAHYKDPFPC 596
DB 537 GOECVEECRVLOGLPREYVNAHCLPCHPECOQNGSVTCFGEADQCACAHYKDPFPC 596
QY 597 VARCPGKVPDLSPYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCAPQASPLTSIVS 656
DB 597 VARCPGKVPDLSPYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCAPQASPLTSIVS 656
QY 657 AVVGILLVVVLGVVFGILIKRQOKIRKYTMRLRLQETELVEPLTPSGAMPNOAQRILK 716
DB 657 AVVGILLVVVLGVVFGILIKRQOKIRKYTMRLRLQETELVEPLTPSGAMPNOAQRILK 716
QY 717 ETELKRVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEIIDEAYVMAG 776
DB 717 ETELKRVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEIIDEAYVMAG 776
QY 777 VGSPYVSRLLGICLTSTVQLTQMLPYGCLLDHVNRGRGLSGDQLLNWCMQIAKMSYL 836
DB 777 VGSPYVSRLLGICLTSTVQLTQMLPYGCLLDHVNRGRGLSGDQLLNWCMQIAKMSYL 836
QY 837 EDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPDKWMALESILR 896
DB 837 EDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPDKWMALESILR 896
QY 897 RBFTHQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPICTIDVYMW 956
DB 897 RBFTHQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPICTIDVYMW 956
QY 957 KCWMIDSECHPRRELVSERMDARDQRFVWIONEDLGPASLDSTFYRSLLDDMDGD 1016

Db 957 KCWMIIDSECRPRELVSFSESRWARDPQREVVIONEDLGASPLDSTFYKSLLEDDMDG 1016
QY 1017 LVDAEEYLVPOQGFPCDPAPGAGGVHHRHRSSTRSGGDLTLGLEPSEEEAPRPLA 1076
Db 1017 LVDAEEYLVPOQGFPCDPAPGAGGVHHRHRSSTRSGGDLTLGLEPSEEEAPRPLA 1076
QY 1077 PSEGASDVDPDGLGWAAGLQSLPTHDPSPLOQRYSEDTVPBPSETDGYVAPLTCSPQ 1136
Db 1077 PSEGASDVDPDGLGWAAGLQSLPTHDPSPLOQRYSEDTVPBPSETDGYVAPLTCSPQ 1136
QY 1137 PEYVNPDPVRPQPPSPREGPLPAARAGATLERAKTILSPCKNGVWVDVAFGAVENPEY 1196
Db 1137 PEYVNPDPVRPQPPSPREGPLPAARAGATLERAKTILSPCKNGVWVDVAFGAVENPEY 1196
QY 1197 LTPQGGAAPOPHPPAFSPAFDNLYWDQPPPERGAPPSTFKGTPTAENPEYLGLOVPV 1255
Db 1197 LTPQGGAAPOPHPPAFSPAFDNLYWDQPPPERGAPPSTFKGTPTAENPEYLGLOVPV 1255

RESULT 8
AAy84780
ID AAY84780 standard; Protein; 1255 AA.
XX
AC AAY84780;
XX
DT 08-AUG-2000 (first entry)
XX
DE Amino acid sequence of the SPLICE erbB-2 receptor protein.
XX
KW SPLICE erbB-2 receptor protein; cell transformation disorder; cancer;
KW tumor cell proliferation; tissue degeneration; arthropathy;
KW bone resorption; inflammatory disease; degenerative disorder;
KW wound healing.
XX
OS Homo sapiens.
XX
PN W0200020579-A1.
XX
PD 13-APR-2000.
XX
PF 01-OCT-1999; 99WO-CA00912.
XX
PR 02-OCT-1998; 98US-0165192.
XX
PA (UYMC-) UNIV MCMASTER.
XX
PI Muller WJ, Siegel PM;
XX
XX WPI; 2000-303768/26.
DR N-PSDB; AAA14812.
XX
XX Nucleic acid encoding an erbB 2 receptor protein designated SPLICE
PT erbB-2, inhibitors of the protein are useful for treatment of cancer -
XX
PS Claim 3; Fig 2; 60pp; English.
XX
XX The present sequence represents a SPLICE erbB-2 receptor protein. The
CC protein has an in-frame deletion of 16 amino acids, 2 of which are
CC conserved cysteine residues, compared to the unspliced protein. The
CC erbB-2 polynucleotide is used to construct probes for detecting
CC disorders of cell transformation such as cancer. Antibodies to the
CC protein may be used to detect SPLICE erbB-2 in a sample. Agents
CC (e.g. antisense oligonucleotides) which inhibit the expression of
CC SPLICE erbB-2 are useful for reducing tumor cell proliferation and
CC treating cancer. Substances which stimulate SPLICE erbB-2 are useful
CC for treating conditions involving damaged cells including conditions
CC in which degeneration of tissue occurs, such as arthropathy, bone
CC resorption, inflammatory diseases, degenerative disorders of the
CC central nervous system and wound healing.
XX
SQ Sequence 1255 AA;

Query Match 98.3%; Score 6694; DB 21; Length 1255;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1237; Conservative 5; Mismatches 9; Indels 8; Gaps 2;
QY 1 MELAALCRWGLLALLPPGAASSTQVCTFNNFTVSFMRVPKVSASHLE---LYQCQCVV 56
Db 1 MELAALCRWGLLALLPPGAASSTQVCTGDMK----LRUPASPETHLDMRLHLYQCQCVV 56
QY 57 QGNLELYLPTNASLSFLQDIOEVQGYVLIHAHQVRQVLPQRLIRIVRGTLQFEDNYALAV 116
Db 57 QGNLELYLPTNASLSFLQDIOEVQGYVLIHAHQVRQVLPQRLIRIVRGTLQFEDNYALAV 116
QY 117 LONGDPLNNTTPTVGTASPGGLRELQRLSLTEILKGVLIQRNPOLCYQDTILWKDIFHN 176
Db 117 LONGDPLNNTTPTVGTASPGGLRELQRLSLTEILKGVLIQRNPOLCYQDTILWKDIFHN 176
QY 177 NOLALTLDITNSRACHPCSPCKGSRGWSESDCQSLTRTVACGACARCKGLPTDCC 236
Db 177 NOLALTLDITNSRACHPCSPCKGSRGWSESDCQSLTRTVACGACARCKGLPTDCC 236
QY 237 HEQCAAGCTGFKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCV 296
Db 237 HEQCAAGCTGFKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCV 296
QY 297 TACPNYLSLTVGSCCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRV 356
Db 297 TACPNYLSLTVGSCCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRV 356
QY 357 TSANTQEFAGCKKIFGSLAFPLPESPDGPASNTAPLOPELOVFFETLEITGYLYISAMP 416
Db 357 TSANTQEFAGCKKIFGSLAFPLPESPDGPASNTAPLOPELOVFFETLEITGYLYISAMP 416
QY 417 DSLPDLVSFQNLQVIRGRILHNGAYSILTQGLGISWLGRLSRLSGSLALHNNHLCF 476
Db 417 DSLPDLVSFQNLQVIRGRILHNGAYSILTQGLGISWLGRLSRLSGSLALHNNHLCF 476
QY 477 VHTVPWDQLFFNPQALHTANRPDECEVGEGLACHQLCARGHCGWPGTQCVCNQFLR 536
Db 477 VHTVPWDQLFFNPQALHTANRPDECEVGEGLACHQLCARGHCGWPGTQCVCNQFLR 536
QY 537 GQECVEECRVLOGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFC 596
Db 537 GQECVEECRVLOGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFC 596
QY 597 VARCPGKVPDLSYMPINWKFDEEGACQPCINCTHSCVDLDDKGPAPQASPLTSIVS 656
Db 597 VARCPGKVPDLSYMPINWKFDEEGACQPCINCTHSCVDLDDKGPAPQASPLTSIVS 656
QY 657 AVVGILLVVVLGVVFGILIKRQOKIRKYTMRLRLOETELVEPLTPSGAMPNQAQMRILK 716
Db 657 AVVGILLVVVLGVVFGILIKRQOKIRKYTMRLRLOETELVEPLTPSGAMPNQAQMRILK 716
QY 717 ETELKVKVLGSGAGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAG 776
Db 717 ETELKVKVLGSGAGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAG 776
QY 777 VGSPVSVRLGLCLTSTVOLTPQLMPYGLLDHVRNRCRLGSDLLNWCMIAGMSYL 836
Db 777 VGSPVSVRLGLCLTSTVOLTPQLMPYGLLDHVRNRCRLGSDLLNWCMIAGMSYL 836
QY 837 EDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGVPIKWMALESILR 896
Db 837 EDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGVPIKWMALESILR 896
QY 897 RPTHQSDVMSYGVTVWELMTGAKPYDGIIPAREIPDLEKGERLPDPPICITDYMIMV 956
Db 897 RPTHQSDVMSYGVTVWELMTGAKPYDGIIPAREIPDLEKGERLPDPPICITDYMIMV 956
QY 957 KCWMIIDSECRPRELVSFSESRWARDPQREVVIONEDLGASPLDSTFYKSLLEDDMDG 1016
Db 957 KCWMIIDSECRPRELVSFSESRWARDPQREVVIONEDLGASPLDSTFYKSLLEDDMDG 1016
QY 1017 LVDAEEYLVPOQGFPCDPAPGAGGVHHRHRSSTRSGGDLTLGLEPSEEEAPRPLA 1076

Db 1137 PEYVQDVRPQPSPREGPLPAARPAAGATLERPKTLSPGKNGVVDVFAFGGAVENPEY 1196
QY 1197 LTPQGGAAPOHPHPAFSPAFDNLYYWDQDPPRGGAPPSTFKGTPTAENPEYLGLDVPV 1255
Db 1197 LTPQGGAAPOHPHPAFSPAFDNLYYWDQDPPRGGAPPSTFKGTPTAENPEYLGLDVPV 1255

RESULT 10

AAG88267
ID AAG88267 standard; Protein; 1255 AA.
XX AAG88267;
XX AC AAG88267;
XX DT 11-SEP-2001 (first entry)
XX HER2/neu amino acid sequence.
XX Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
KW immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;
KW tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
XX Homo sapiens.
XX WO200141787-A1.
XX 14-JUN-2001.
XX 11-DEC-2000; 2000WO-US33591.
XX 10-DEC-1999; 99US-0458299.
XX (EPIM-) EPIMMUNE INC.
XX Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;
PI Keogh E;
XX WPI; 2001-374995/39.

DR An isolated prepared HER2/neu epitope useful in a vaccine for inducing
PT cellular immune responses for the prevention and treatment of cancer -
XX Disclosure; Page 15; 199pp; English.
XX The present invention describes isolated prepared HER2/neu epitopes (I).
CC Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
CC culture in vitro and binds to a complex of an epitope (I), bound to a
CC human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)
CC and a second epitope and the peptide is less than 50 contiguous amino
CC acids that have 100% identity with a native peptide sequence of HER2/neu;
CC (3) a vaccine composition (III) comprising (II) and a pharmaceutical
CC excipient; (4) an isolated nucleic acid encoding a peptide comprising
CC (I); and (5) an isolated nucleic acid encoding (II). (I) has cytostatic
CC and immunostimulant activities, and can be used in vaccines. (I), (II)
CC and (III) are useful for inducing cellular immune responses for the
CC prevention and treatment of cancer. (I) and (II) are useful for
CC monitoring or evaluating an immune response to a tumour-associated
CC antigen when incubated with a T lymphocyte sample from a patient and
CC detecting the presence of bound T lymphocyte to (I) or (II). Epitope
CC based vaccines mean that immunosuppressive epitopes that may be present
CC in whole antigens may be avoided. Selected epitopes may be combined to
CC enhance immunogenicity. The possible pathological side effects caused by
CC infectious agents or whole protein antigen is eliminated. The vaccine
CC provides the ability to direct and focus an immune response to multiple
CC selected antigens from the same pathogen. Epitope-based anti-tumour
CC vaccines provides the opportunity to combine epitopes derived from
CC multiple tumour-associated molecules addressing the problem of tumour-
CC tumour variability and reducing the likelihood of tumour 'escape' due to
CC antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
CC the exemplification of the present invention.

XX Sequence 1255 AA;
SQ Query Match 98.3%; Score 6694; DB 22; Length 1255;

Best Local Similarity 98.3%; Pred. No. 0;
Matches 1237; Conservative 5; Mismatches 9; Indels 8; Gaps 2;
QY 1 MELAALCRWGLLALLPPGAASSTQVCTFNNFVSEFWLRVPKVSASHLE----LYQCQGVV 56
Db 1 MELAALCRWGLLALLPPGAASSTQVCTGDMK----LRLPASPEHDLMLRHLHYQCQGVV 56
QY 57 QGNLELTYPNTNASISFLQDIOEVQGYVLIANQVRQVPLQRLRIVRGTQOLFEDNYALAV 116
Db 57 QGNLELTYPNTNASISFLQDIOEVQGYVLIANQVRQVPLQRLRIVRGTQOLFEDNYALAV 116
QY 117 LDNGDPLNNTTPTVTCASPGGIRELQRLSITELKGGVLIQRPOLCYQDTILWKDIFHKN 176
Db 117 LDNGDPLNNTTPTVTCASPGGIRELQRLSITELKGGVLIQRPOLCYQDTILWKDIFHKN 176
QY 177 NOLALTLIDTNRSRACHPCSMCKSGRCWGESSEDCQSLTRTVTCAGGCARCKGPLETDC 236
Db 177 NOLALTLIDTNRSRACHPCSMCKSGRCWGESSEDCQSLTRTVTCAGGCARCKGPLETDC 236
QY 237 HEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCV 296
Db 237 HEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCV 296
QY 297 TACPYNILSTDVSGCTLVCPHNOEVTAEADGTQRCCKSPCARVCYGLGMEHLREVRVAV 356
Db 297 TACPYNILSTDVSGCTLVCPHNOEVTAEADGTQRCCKSPCARVCYGLGMEHLREVRVAV 356
QY 357 TSANIQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFPETLEITGVLYISAMP 416
Db 357 TSANIQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFPETLEITGVLYISAMP 416
QY 417 DSLPDLISVFQNLQVIRGRILHNGAYSILTLQGLGISWLGSLRLSRELGLALIHNTLHLCF 476
Db 417 DSLPDLISVFQNLQVIRGRILHNGAYSILTLQGLGISWLGSLRLSRELGLALIHNTLHLCF 476
QY 477 VHTVWDQLFRPHQALLHTANRPEDECVGEGACHOLCARGHCWGPGTQCVCNSQFLR 536
Db 477 VHTVWDQLFRPHQALLHTANRPEDECVGEGACHOLCARGHCWGPGTQCVCNSQFLR 536
QY 537 GQECVEECKVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPKPPC 596
Db 537 GQECVEECKVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPKPPC 596
QY 597 VARCPGKVKPELSYMPIWKFPDEEGACQPCINCTHSCVDLDDKGPABQASPLTSIVS 656
Db 597 VARCPGKVKPELSYMPIWKFPDEEGACQPCINCTHSCVDLDDKGPABQASPLTSIVS 656
QY 657 AVUGILLVVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILK 716
Db 657 AVUGILLVVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILK 716
QY 717 ETELRRKVKVLSGAFCTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEIILDEAYVMAG 776
Db 717 ETELRRKVKVLSGAFCTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEIILDEAYVMAG 776
QY 777 VGSPPVSRLLGLCLTSTVOLVTLQMPYGCLLDHVNRGRGLSQDLNLMQIAKMSYL 836
Db 777 VGSPPVSRLLGLCLTSTVOLVTLQMPYGCLLDHVNRGRGLSQDLNLMQIAKMSYL 836
QY 837 EDVRLVHRDLAARNLVKSPNHVKITDFGLABLLIDETEHADGGKVPKIMWALESIILR 896
Db 837 EDVRLVHRDLAARNLVKSPNHVKITDFGLABLLIDETEHADGGKVPKIMWALESIILR 896
QY 897 RFRTHQSDVMSYGVVTWELMTGAKPYDGI PAREIPDLLEKGERLPPOPICTIDVYIMV 956
Db 897 RFRTHQSDVMSYGVVTWELMTGAKPYDGI PAREIPDLLEKGERLPPOPICTIDVYIMV 956
QY 957 KCWMIIDSECRPRFRELVSFSESMARDPQRFVVIQNEIDLGPASPLDSTFVRSLEDDMDGD 1016
Db 957 KCWMIIDSECRPRFRELVSFSESMARDPQRFVVIQNEIDLGPASPLDSTFVRSLEDDMDGD 1016
QY 1017 LVDAAEYLVPOQGFPCPDPAAGVMVHRHRSSTRSGGDLTLGLPSEEPSPPLA 1076
Db 1017 LVDAAEYLVPOQGFPCPDPAAGVMVHRHRSSTRSGGDLTLGLPSEEPSPPLA 1076

Db	1017	LVDABEYLVPQQGFCDDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEBAPRSPLA	1076
QY	1077	PSEGAGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQ	1136
Db	1077	PSEGAGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQ	1136
QY	1137	PEYVQPDVVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVYKDVFAFGGAVENPEY	1196
Db	1137	PEYVQPDVVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVYKDVFAFGGAVENPEY	1196
QY	1197	LTPQGAAPQHPHPPAPSPAFDNLYYWDQPPPERGAPSTFKGTPTAENPEYILGLDVPV	1255
Db	1197	LTPQGAAPQHPHPPAPSPAFDNLYYWDQPPPERGAPSTFKGTPTAENPEYILGLDVPV	1255
RESULT 11			
AAE24067			
ID	AAE24067 standard; Protein; 1255 AA.		
AC	AAE24067;		
XX			
DT	23-SEP-2002 (first entry)		
XX			
DE	Human Her-2 protein.		
XX			
KW	Human; Her-2; epidermal growth factor receptor 2; infection; cancer;		
KW	hyperproliferative disorder; prophylaxis; inflammation; antisense;		
KW	tumour; gene therapy; phosphorothioate backbone.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200222636-A1.		
XX			
PD	21-MAR-2002.		
XX			
PF	12-SEP-2001; 2001WO-US28572.		
XX			
PR	15-SEP-2000; 2000US-0663834.		
XX			
PA	(ISIS-) ISIS PHARM INC.		
XX			
PI	Bennett CF, Cowser LM;		
XX			
DR	WPI: 2002-471192/50.		
DR	N-PSDB; AAD38904.		
XX			
PT	Novel antisense oligonucleotide which modulates the expression of Human		
PT	Epidermal Growth Factor receptor, Her2, is useful for treating tumors		
PT	inflammation or to prevent infection in humans -		
XX			
PS	Example 13; Page 95-107; 116pp; English.		
XX			
CC	The invention relates to antisense compounds targetted to a nucleic		
CC	acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)		
CC	that specifically hybridises with and inhibits the expression of Her2.		
CC	Antisense compounds of the invention are used for treating diseases or		
CC	conditions associated with Her2 such as hyperproliferative disorders		
CC	e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,		
CC	neural or cardiac cancer. They are also useful prophylactically e.g.		
CC	to prevent or delay infection, inflammation and tumour formation. The		
CC	invention is also used in gene therapy. The present sequence is human		
CC	Her-2 protein.		
XX			
SQ	Sequence 1255 AA;		
Query Match			
Best Local Similarity 98.3%; Score 6694; DB 23; Length 1255;			
Matches 1237; Conservative 5; Mismatches 9; Indels 8; Gaps 2;			
QY	1	MELAAACRWGLLLALPPGGAASQVCTFNFTVSWLRVPKVSASHLE----	LYQGCQVV 56
Db	1	MELAAACRWGLLLALPPGGAASQVCTGDMK-----LRUPASPETHLDMLRHLYQGCQVV	56

QY	57	QGNLELYLPTNASLSFLQDIOEVQGYVLIHAHQVRQVPLQRLIRIVRGTLQEDFNALAV	1116
Db	57	QGNLELYLPTNASLSFLQDIOEVQGYVLIHAHQVRQVPLQRLIRIVRGTLQEDFNALAV	1116
QY	117	LNGDPLNNTTPTVGTASPGGLRELQRLSUTELKGGVLIQRNPOLCYQDTILWKDIFHKN	176
Db	117	LNGDPLNNTTPTVGTASPGGLRELQRLSUTELKGGVLIQRNPOLCYQDTILWKDIFHKN	176
QY	177	NOLALTLIDTNSRACHPCSPMKGSRGWESSEDCQSLTRTVAGGCARCKGLPTDCC	236
Db	177	NOLALTLIDTNSRACHPCSPMKGSRGWESSEDCQSLTRTVAGGCARCKGLPTDCC	236
QY	237	HEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCV	296
Db	237	HEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCV	296
QY	297	TACPNYLTSDVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMEHLRVRVAV	356
Db	297	TACPNYLTSDVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMEHLRVRVAV	356
QY	357	TSANTOEFAGCKKIFGSLAFLPESPDGPASNTAPLQPELOVFEETLEITGYLYISAMP	416
Db	357	TSANTOEFAGCKKIFGSLAFLPESPDGPASNTAPLQPELOVFEETLEITGYLYISAMP	416
QY	417	DSLPLDSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLSGSLALIHNTLCLF	476
Db	417	DSLPLDSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLSGSLALIHNTLCLF	476
QY	477	VHTVPWDQLFRNPHOALLHTANRPDECEVGEGLACHQLCARGHCWGPGTQCVCNQFLR	536
Db	477	VHTVPWDQLFRNPHOALLHTANRPDECEVGEGLACHQLCARGHCWGPGTQCVCNQFLR	536
QY	537	GOECVEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDPFPC	596
Db	537	GOECVEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDPFPC	596
QY	597	VARCPGKVPDLSYMPKFWPDEEGACQPCPINCTHSCVDLDDKGPAPQASPLTSIVS	656
Db	597	VARCPGKVPDLSYMPKFWPDEEGACQPCPINCTHSCVDLDDKGPAPQASPLTSIVS	656
QY	657	AVVGIILLVVVLGVVFGILIKRRQKIRKYTMRLLOETELVEPLTPSGAMPNQAQRIILK	716
Db	657	AVVGIILLVVVLGVVFGILIKRRQKIRKYTMRLLOETELVEPLTPSGAMPNQAQRIILK	716
QY	717	ETELRKVKVLGSGAGCTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAVVMAG	776
Db	717	ETELRKVKVLGSGAGCTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAVVMAG	776
QY	777	VGSPYVSRLLGICLTSTVOLVTQLMPYGCLLDHVRENRLGSLQDILLNMQIAGMSYL	836
Db	777	VGSPYVSRLLGICLTSTVOLVTQLMPYGCLLDHVRENRLGSLQDILLNMQIAGMSYL	836
QY	837	EDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPDKWMALESILR	896
Db	837	EDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPDKWMALESILR	896
QY	897	RSFTQSDVWSGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPDPICTIDVYIMV	956
Db	897	RSFTQSDVWSGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPDPICTIDVYIMV	956
QY	957	KCWMIDSECRPRFRELVSFSEFSEFSEFSEFSEFSEFSEFSEFSEFSEFSEFSEFSEF	1016
Db	957	KCWMIDSECRPRFRELVSFSEFSEFSEFSEFSEFSEFSEFSEFSEFSEFSEFSEF	1016
QY	1017	LVDABEYLVPQQGFCDDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEBAPRSPLA	1076
Db	1017	LVDABEYLVPQQGFCDDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEBAPRSPLA	1076
QY	1077	PSEGAGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQ	1136
Db	1077	PSEGAGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQ	1136
QY	1137	PEYVQPDVVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVYKDVFAFGGAVENPEY	1196

Db 1137 PEYVQDFVRPQPPSPREGPLPAARAGATLERPKTLSPGKNGVVKDVFAFGAVENPEY 1196
QY 1197 LTPQGGAAQPPHPPAFSPAFDNLXYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255
Db 1197 LTPQGGAAQPPHPPAFSPAFDNLXYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255
RESULT 12
AAE20479
ID AAE20479 standard; Protein; 1255 AA.
XX
AC AAE20479;
XX
DT 01-JUL-2002 (first entry)
XX
DE Human Her-2/neu protein.
XX
KW Human; Her-2/Neu protein; immune response; gene therapy; breast cancer;
KW human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Region 1021..1030
FT /note= "Naturally processed HLA-B44-restricted epitope"
XX
PN WO200214503-A2.
XX
PD 21-FEB-2002.
XX
PF 14-AUG-2001; 2001WO-US41733.
XX
PR 14-AUG-2000; 2000US-225152P.
PR 28-SEP-2000; 2000US-236428P.
PR 21-FEB-2001; 2001US-270520P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Hand-zimmermann S, Cheever MA, Foy TW, Lodes MJ, Kalos MD;
PI McNeill PD, Vedvick TS;
XX
DR WPI; 2002-280758/32.
DR N-PSDB; AAD32743.
XX
PT Novel isolated Her-2/Neu polypeptide composition useful for therapy,
PT prevention and diagnosis of cancer, preferably breast cancer
XX
PS Disclosure; Page 114-117; 129pp; English.
XX
CC The invention relates to an isolated Her-2/Neu polypeptide composition
CC effective for eliciting an immune response. The invention is useful for
CC eliciting an immune response in a patient, where the patient is human
CC leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.
CC The composition is useful for the therapy and diagnosis of cancer,
CC preferably breast cancer, in pharmaceutical compositions, e.g., vaccine
CC and other compositions for the diagnosis, prevention and treatment of
CC human malignancies, for stimulating and/or expanding T cells specific for
CC Her-2/Neu polypeptide and for inhibiting the development of cancer in a
CC patient. The invention is useful for stimulating a T cell response in a
CC human patient, as probe or primer for nucleic acid hybridisation, to
CC selectively form duplex molecules with complementary stretches of the
CC entire Her-2/Neu gene or gene fragments of interest, to isolate a full
CC length gene from a suitable library, and to direct expression of a
CC polypeptide in appropriate host cells. The composition is useful in
CC prophylactic or therapeutic applications and for the treatment of cancer,
CC preferably for the immunotherapy of breast cancer and other Her-2/Neu-
CC associated malignancies. The invention is useful in gene therapy. The
CC present sequence is human Her-2/neu protein.
XX
SQ Sequence 1255 AA;

Query Match 98.3%; Score 6694; DB 23; Length 1255;

Best Local Similarity 98.3%; Pred. No. 0;
Matches 1237; Conservative 5; Mismatches 9; Indels 8; Gaps 2;
QY 1 MELAALCRWGLLALLPPGAASTQVCTFNNFVSVFWLRVPKVSASHLE----LYQCQCVV 56
Db 1 MELAALCRWGLLALLPPGAASTQVCTGDMK----LRLPASPETHDLMLRHLHYQCQCVV 56
QY 57 QGNLELTYPNVASISFLQDIOEVQGYVLIJAHNQVRQVPLQRLRIVRGTLQEDFNVALAV 116
Db 57 QGNLELTYPNVASISFLQDIOEVQGYVLIJAHNQVRQVPLQRLRIVRGTLQEDFNVALAV 116
QY 117 LDNGDPLANNTPVTGASPGGLRELOLRSITELTKGGVLIQRPOLCYQDTILWKDLFHKN 176
Db 117 LDNGDPLANNTPVTGASPGGLRELOLRSITELTKGGVLIQRPOLCYQDTILWKDLFHKN 176
QY 177 NQLALTLIDTNRSRACHPCSPMKGSRGWGESSEDCQSLTRTVCAAGCARCKGPLPTDCC 236
Db 177 NQLALTLIDTNRSRACHPCSPMKGSRGWGESSEDCQSLTRTVCAAGCARCKGPLPTDCC 236
QY 237 HEQCAAGCTGPIKHSDDLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCV 296
Db 237 HEQCAAGCTGPIKHSDDLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCV 296
QY 297 TACPYNYLSTDVGSCTLYCPLHNOEVTAEADGTORCEKSKPCARVCYGLGMEHLREVAV 356
Db 297 TACPYNYLSTDVGSCTLYCPLHNOEVTAEADGTORCEKSKPCARVCYGLGMEHLREVAV 356
QY 357 TSANIQEFAGCCKIFGSLAFIPESFDGPPASNTAPLQPEQLQVFTLEITGYLISAMP 416
Db 357 TSANIQEFAGCCKIFGSLAFIPESFDGPPASNTAPLQPEQLQVFTLEITGYLISAMP 416
QY 417 DSLPDLVSFQNLQVIRGRILHNGAYSLTLQGLIGISWGLRSURELGSGIALIHHNTHLCF 476
Db 417 DSLPDLVSFQNLQVIRGRILHNGAYSLTLQGLIGISWGLRSURELGSGIALIHHNTHLCF 476
QY 477 VHTVPDQLFRPHQALLHTANRPEDECVGEGLAGHOLCARGCHGWPGTQCVNCSQFLR 536
Db 477 VHTVPDQLFRPHQALLHTANRPEDECVGEGLAGHOLCARGCHGWPGTQCVNCSQFLR 536
QY 537 GOECVEECRVLOGLPREVYNARHCLPCHPECPQNGSVTCFPGPADQCVACAHYKDPFPC 596
Db 537 GOECVEECRVLOGLPREVYNARHCLPCHPECPQNGSVTCFPGPADQCVACAHYKDPFPC 596
QY 597 VARCPGKVPDLVSYPIMKFPDEEGACQPCPINCTHSCVDLDDKGCAPQASPLTSIVS 656
Db 597 VARCPGKVPDLVSYPIMKFPDEEGACQPCPINCTHSCVDLDDKGCAPQASPLTSIVS 656
QY 657 AVVGILLVVVLGVWFGILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAQMRILK 716
Db 657 AVVGILLVVVLGVWFGILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAQMRILK 716
QY 717 ETELKRVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEIIDEAYVMAG 776
Db 717 ETELKRVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEIIDEAYVMAG 776
QY 777 VGSPVSVRLIGLICLTSTVQLVTQLMPYGLLDHVRENRRGLSGQDLINWCMQIAKMSYL 836
Db 777 VGSPVSVRLIGLICLTSTVQLVTQLMPYGLLDHVRENRRGLSGQDLINWCMQIAKMSYL 836
QY 837 EDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMWALESIUR 896
Db 837 EDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMWALESIUR 896
QY 897 RPFTHQSDVWSGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPPICITDVMYMW 956
Db 897 RPFTHQSDVWSGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPPICITDVMYMW 956
QY 957 KCWMIIDSECRPFRFELVSEFSRMARDPQRFVVIQNEDELGPASPLDSTFYRSLLEDDMGD 1016
Db 957 KCWMIIDSECRPFRFELVSEFSRMARDPQRFVVIQNEDELGPASPLDSTFYRSLLEDDMGD 1016
QY 1017 LVDASEYILVPQOGFFCPDPAAGAGMWVHHRSSSTRSGGDLTLGLSPSESEAPRPLA 1076
Db 1017 LVDASEYILVPQOGFFCPDPAAGAGMWVHHRSSSTRSGGDLTLGLSPSESEAPRPLA 1076

Db 1017 LVDAEYLVQOQFCFDPAPAGAGVHHRSSSTRSGGDLTLGLEPSEEARPSPLA 1076
QY 1077 PSEGAGSDVFDGLGMAAGLQSLPHDPSPLQRYSEDTVPLPSETDCGYVAPLTCSPQ 1136
Db 1077 PSEGAGSDVFDGLGMAAGLQSLPHDPSPLQRYSEDTVPLPSETDCGYVAPLTCSPQ 1136
QY 1137 PEYVNPQDVPRQPPSPREGPLPAARPAAGATLERAKTLPSPKNGVVKDVFAFGGAVENPEY 1196
Db 1137 PEYVNPQDVPRQPPSPREGPLPAARPAAGATLERAKTLPSPKNGVVKDVFAFGGAVENPEY 1196
QY 1197 LTPQGAAPQHPHPPAFSPAFDNLVYWDQPPPERGAPPSTFKGTPAENPEYLGLDVVP 1255
Db 1197 LTPQGAAPQHPHPPAFSPAFDNLVYWDQPPPERGAPPSTFKGTPAENPEYLGLDVVP 1255

RESULT 13

AA51143
ID AA51143 standard; Protein; 1255 AA.
AC AA51143;
XX 17-JUN-2002 (first entry)
XX Human Her-2/neu oncogene-encoded p185 glycoprotein.
DE Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;
KW tyrosine kinase; receptor; c-erbB2; gene therapy.
XX Homo sapiens.
XX Key Location/Qualifiers
FH Domain 1..653
FT Domain /note= "extracellular domain"
FT Domain 676..1255
FT Domain /note= "intracellular domain"
FT Domain 990..1255
FT FT /note= "phosphorylation domain"
XX WO200212341-A2.
XX
XX
XX 14-FEB-2002.
XX 03-AUG-2001; 2001WO-US24283.
XX
XX 03-AUG-2000; 2000US-0632507.
XX (CORI-) CORIXA CORP.
XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX Cheever MA, Gheysen D;
XX WPI; 2002-241743/29.
XX N-PSDB; ABA92250.
XX
XX Her-2/neu fusion protein for treating or preventing cancer by eliciting
XX or enhancing an immune response to the protein, has Her-2/neu
XX extracellular domain fused to Her-2/neu intracellular or
XX phosphorylation domain -
XX Claim 68; Fig 7; 141pp; English.

XX The present sequence is that of human Her-2/neu (p185 glycoprotein
XX or c-erbB2), an oncogenic self-protein and target for anti-cancer
XX vaccines. The Her-2/neu gene is amplified and p185 is overexpressed
XX in a variety of cancers, including breast, ovarian, colon, lung and
XX prostate cancer. Her-2/neu is a member of the tyrosine kinase
XX family of receptor-like glycoproteins. It comprises an extracellular
XX domain with homology to the epidermal growth factor receptor
XX (EGFR), a highly hydrophobic transmembrane domain and a C-terminal
XX intracellular domain that also shows homology to EGFR. Its
XX overexpression correlates with a poor prognosis in breast and
XX ovarian cancers. The invention provides Her-2/neu fusion
XX proteins, nucleic acids encoding them, viral vectors, and vaccines

CC comprising the fusion proteins or nucleic acid molecules. In
CC preferred fusion proteins, the extracellular domain of a Her-2/neu
CC protein is fused to a Her-2/neu intracellular domain or
CC phosphorylation domain (or its DeltapP fragment). An immune
CC response to Her-2/neu protein is elicited or enhanced by
CC administering the fusion protein in the form of a vaccine, or by
CC transfecting cells of an animal ex vivo with a nucleic acid
CC encoding the fusion protein, and delivering the transfected cells
CC to the animal. The fusion proteins, nucleic acids, and isolated
CC specific T-cells are useful for inhibiting the development of a
CC cancer, especially breast, ovarian, colon, lung or prostate cancer
CC in a patient. T cells that specifically react with a Her-2/neu
CC fusion protein can be used to remove tumour cells from a sample in
CC order to inhibit the development of cancer in a patient.

XX Sequence 1255 AA;

Query Match 98.3%; Score 6694; DB 23; Length 1255;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1237; Conservative 5; Mismatches 9; Indels 8; Gaps 2;
QY 1 MELAALCRWGLLLALLPPGAASTQVCTFNNFTVSFWLRVPKVSASHLE----LYQCQCVV 56
Db 1 MELAALCRWGLLLALLPPGAASTQVCTGDMK----LRLPASPETHLDMRLHLYQCQCVV 56
QY 57 QGNLELTYPNTNASLSFLQDIOEVQGYVLIANQVQVPLQRLIRVRGTQLPEDNVALAV 116
Db 57 QGNLELTYPNTNASLSFLQDIOEVQGYVLIANQVQVPLQRLIRVRGTQLPEDNVALAV 116
QY 117 LDNGDPLNNTTPTVTGASPGGLRELQRLSLTEILKGVLIQRPOLCYQDTILWKDIFHKN 176
Db 117 LDNGDPLNNTTPTVTGASPGGLRELQRLSLTEILKGVLIQRPOLCYQDTILWKDIFHKN 176
QY 177 NQALTLDTNRSRACHPCSPCKSRGWGESSEDCQSLTRTVCCAGCARCGPLPTDCC 236
Db 177 NQALTLDTNRSRACHPCSPCKSRGWGESSEDCQSLTRTVCCAGCARCGPLPTDCC 236
QY 237 HEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESNPNEGRTYTCASCV 296
Db 237 HEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESNPNEGRTYTCASCV 296
QY 297 TACPYNLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRAV 356
Db 297 TACPYNLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRAV 356
QY 357 TSANTQEFAGCKKIFGSLAFLESFDGDPASNTAPLQPEQLQVFETLEITGYLYISAWP 416
Db 357 TSANTQEFAGCKKIFGSLAFLESFDGDPASNTAPLQPEQLQVFETLEITGYLYISAWP 416
QY 417 DSLPDLVSFQNLQVIRGRILHNGAYSILTQGLGISWLGRLSRLSGSLALHNNTHLCF 476
Db 417 DSLPDLVSFQNLQVIRGRILHNGAYSILTQGLGISWLGRLSRLSGSLALHNNTHLCF 476
QY 477 VHTVPMQDLFRNPQALLHTANRPEDECYGEGLACHQLCARGCHGPGTQCVCNCSQFLR 536
Db 477 VHTVPMQDLFRNPQALLHTANRPEDECYGEGLACHQLCARGCHGPGTQCVCNCSQFLR 536
QY 537 GOECVBECEKRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPC 596
Db 537 GOECVBECEKRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPC 596
QY 597 VARPSPGVKPDLSYMPINWKFDEEGACOPCINCTHSCVDLDDKGPAPORASPLTSIVS 656
Db 597 VARPSPGVKPDLSYMPINWKFDEEGACOPCINCTHSCVDLDDKGPAPORASPLTSIVS 656
QY 657 AVVGILLVVVLGVWFVGLIKRQQKIRKYTWRLLEQETELVEPLTPSGAMPNQAMRILK 716
Db 657 AVVGILLVVVLGVWFVGLIKRQQKIRKYTWRLLEQETELVEPLTPSGAMPNQAMRILK 716
QY 717 ETELKRVKVLGSGAGTGYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAG 776
Db 717 ETELKRVKVLGSGAGTGYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAG 776

QY 777 VGSPPYVSRLLGICLTSTVQLVTOLMPYGCGLLDHVRENHRLGSGDILLNWCQIAKMSYL 836
Db 777 VGSPPYVSRLLGICLTSTVQLVTOLMPYGCGLLDHVRENHRLGSGDILLNWCQIAKMSYL 836
QY 837 EDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPPIKWMALLESILR 896
Db 837 EDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPPIKWMALLESILR 896
QY 897 RRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICTIDVYMIW 956
Db 897 RRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICTIDVYMIW 956
QY 957 KCMWIDSECPRELVSEFSRMARDQRFVVIQNEIDLGPASPLDSTFVRSILLEDDMGD 1016
Db 957 KCMWIDSECPRELVSEFSRMARDQRFVVIQNEIDLGPASPLDSTFVRSILLEDDMGD 1016
QY 1017 LVDAEYLVPOQGFCDPAPGAGGMVHHRHRSSTRSGGDLTLGLEPSEEAAPSPPLA 1076
Db 1017 LVDAEYLVPOQGFCDPAPGAGGMVHHRHRSSTRSGGDLTLGLEPSEEAAPSPPLA 1076
QY 1077 PSEGAGSDVFDGLGMAAGLQSLPTHDPSPQLQYSEDPTVPLPSETDGYVAPLTCSPQ 1136
Db 1077 PSEGAGSDVFDGLGMAAGLQSLPTHDPSPQLQYSEDPTVPLPSETDGYVAPLTCSPQ 1136
QY 1137 PEYVNOQDVRPQPPSPREGPLPAARPAGATLERAKTILSPKNGVWVDVPAFGAVENPEY 1196
Db 1137 PEYVNOQDVRPQPPSPREGPLPAARPAGATLERAKTILSPKNGVWVDVPAFGAVENPEY 1196
QY 1197 LTFQGGAAPOPHPPAFSPAFDNLYYWDQDPPPERGAPPSTFKGTPTAENPEYLGLOVPV 1255
Db 1197 LTFQGGAAPOPHPPAFSPAFDNLYYWDQDPPPERGAPPSTFKGTPTAENPEYLGLOVPV 1255

RESULT 14

AAU77114
ID AAU77114 standard; Protein; 1255 AA.
XX
AC AAU77114;
XX
DT 05-JUN-2002 (first entry)
XX
DE Human Her-2/neu polypeptide.
XX
KW Human; Her-2/neu; cytostatic; haematological malignancy; CML;
KW acute myelogenous leukaemia; AMI; chronic myelogenous leukaemia; CLL;
KW chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;
KW Hodgkin's lymphoma; T cell therapy.
XX
OS Homo sapiens.
XX
PN WO200213847-A2.
XX
PD 21-FEB-2002.
XX
PF 13-AUG-2001; 2001WO-US25408.
XX
PR 14-AUG-2000; 2000US-0638280.
PR 28-SEP-2000; 2000US-0675904.
XX
XX
PA (CORI-) CORIXA CORP.
XX
XX
XX Gaiger A, Cheever MA, Hand-zimmermann S;
PI WPI; 2002-280741/32.
DR N-PSDB; ABK10730.
DR
XX
XX Inhibiting haematological malignancy development by administering
PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
PT encoding the polypeptide, or antigen presenting cells expressing the
PT polypeptide
XX
XX Disclosure; Page 71-74; 74pp; English.
XX

CC The invention relates to a method for inhibiting development of
CC haematological malignancy in a patient by administering a polypeptide
CC comprising an immunogenic portion of Her-2/neu or a polynucleotide
CC encoding the polypeptide. Antigen presenting cells that express the
CC protein can also be administered. The sequences are used for inhibiting
CC development of haematological malignancy such as acute myelogenous
CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
CC lymphoma. This sequence represents the human Her-2/neu polypeptide.
XX
SQ Sequence 1255 AA;

Query Match 98.3%; Score 6694; DB 23; Length 1255;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1237; Conservative 5; Mismatches 9; Indels 8; Gaps 2;

QY 1 MELAALCRWGLLLALLPPGAASTVCTFNNFTVSFWLRVPKVSASHLE---LYQSCQVV 56
Db 1 MELAALCRWGLLLALLPPGAASTVCTGTDK-----LRUPASPETHLDMRLHLYQSCQVV 56

QY 57 QGNLELYLPTNASLSFLQDIOEVQGYVLIHAHQVRQVPLQRLIRIVRGTLQFEDNYALAV 116
Db 57 QGNLELYLPTNASLSFLQDIOEVQGYVLIHAHQVRQVPLQRLIRIVRGTLQFEDNYALAV 116

QY 117 LONGDPLNNTTVPVTGASPGGLRELQRLSUTLILKGVLIQRPOLCYQDTILWKDIFHKN 176
Db 117 LONGDPLNNTTVPVTGASPGGLRELQRLSUTLILKGVLIQRPOLCYQDTILWKDIFHKN 176

QY 177 NOLALTLIDTNSRACHPCSPKSGRCWGESSEDCOSLTRTVCAGGCARCKGLPTDCC 236
Db 177 NOLALTLIDTNSRACHPCSPKSGRCWGESSEDCOSLTRTVCAGGCARCKGLPTDCC 236

QY 237 HEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNEPGRYTFGASCV 296
Db 237 HEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNEPGRYTFGASCV 296

QY 297 TACPNYLTSTVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMBHLREVAV 356
Db 297 TACPNYLTSTVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMBHLREVAV 356

QY 357 TSANTQEFAGCKKIFGSLAFLESFDGDPASNTAPLOPELOVFETLEITGYLISAMP 416
Db 357 TSANTQEFAGCKKIFGSLAFLESFDGDPASNTAPLOPELOVFETLEITGYLISAMP 416

QY 417 DSLPDLSPVQNLQVIRGRILHNGAYSILTQGLIGISWLGRLSRLSGSLALIHNTHLCF 476
Db 417 DSLPDLSPVQNLQVIRGRILHNGAYSILTQGLIGISWLGRLSRLSGSLALIHNTHLCF 476

QY 477 VHTVPMDQLFFNPHQALLHTANRPEDECYEGELACHQLCARGHCWGPGPTQCVCNCSQFLR 536
Db 477 VHTVPMDQLFFNPHQALLHTANRPEDECYEGELACHQLCARGHCWGPGPTQCVCNCSQFLR 536

QY 537 GOECVEECRVLOGLPREYNARHCLPCHPECPONGSVTCFGEADOCVACAHYKDPFPC 596
Db 537 GOECVEECRVLOGLPREYNARHCLPCHPECPONGSVTCFGEADOCVACAHYKDPFPC 596

QY 597 VARCPGKVPULSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGPAPQASPLTSIVS 656
Db 597 VARCPGKVPULSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGPAPQASPLTSIVS 656

QY 657 AVVGILLVVVLGVVFGILIKRQOKIRKYTMRLRLLQETELVEPLTPSGAMPNQAQMRILK 716
Db 657 AVVGILLVVVLGVVFGILIKRQOKIRKYTMRLRLLQETELVEPLTPSGAMPNQAQMRILK 716

QY 717 ETELKRVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEIIDEAYVMAG 776
Db 717 ETELKRVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEIIDEAYVMAG 776

QY 777 VGSPVYSRLIGHCLTSTVOLVQLMPYGCGLLDHVRENHRLGSGDILLNWCQIAKMSYL 836
Db 777 VGSPVYSRLIGHCLTSTVOLVQLMPYGCGLLDHVRENHRLGSGDILLNWCQIAKMSYL 836

QY 837 EDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPPIKWMALLESILR 896

Db 837 EDVRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMWALSILR 896
QY 897 RRTTHOSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPCTIDVYMIW 956
Db 897 RRTTHOSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPCTIDVYMIW 956
QY 957 KCMWIDSECRPRFRELVSERFARMARDPQRFVVIQNEEDLGSPASPLDSTFYRSLEDDDDMGD 1016
Db 957 KCMWIDSECRPRFRELVSERFARMARDPQRFVVIQNEEDLGSPASPLDSTFYRSLEDDDDMGD 1016
QY 1017 LVDABEYLVPQGGFFCPDPAPGAGGVVHHRHSSSTRSGGDLTLGLSESEBAPRSPLA 1076
Db 1017 LVDABEYLVPQGGFFCPDPAPGAGGVVHHRHSSSTRSGGDLTLGLSESEBAPRSPLA 1076
QY 1077 PSEGAGSDVFDGLGMAAGKLOSLPHDPSLQRYSEDTVPLPSETDGYVAPLTCSPQ 1136
Db 1077 PSEGAGSDVFDGLGMAAGKLOSLPHDPSLQRYSEDTVPLPSETDGYVAPLTCSPQ 1136
QY 1137 PEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVWVDVAFGGAVENPEY 1196
Db 1137 PEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVWVDVAFGGAVENPEY 1196
QY 1197 LTPQGGAAQPHPPPPAFSPAFDNLVYWDQPPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
Db 1197 LTPQGGAAQPHPPPPAFSPAFDNLVYWDQPPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255

RESULT 15

AAR39568
ID AAR39568 standard; Protein; 1433 AA.

XX AAR39568;

AC AAR39568;
DT 07-FEB-1994 (first entry)

DE Sequence of c-erbB-2 tumour antigen.

XX Tumour antigen; c-erbB-2; glycoprotein.

XX Homo sapiens.

XX WO9316185-A.

XX 19-AUG-1993.

XX 05-FEB-1993; 93WO-US01055.

XX 06-FEB-1992; 92US-0831967.

XX (CETU) CETUS ONCOLOGY CORP.

XX (CREA-) CREATIVE BIOMOLECULES INC.

XX Houston LL, Huston JS, Oppermann H, Ring DB;

XX WPI; 1993-272889/34.

XX N-PSDB; AAQ46083.

XX New single chain Fv polypeptide binding to C-erbB-2 tumour antigen - for imaging or treating breast or ovarian cancer etc.

XX Disclosure; pages 48-54; 87pp; English.

XX c-erbB-2 refers to a protein antigen expressed on the surface of tumour cells, such as breast and ovarian tumour cells, which is an approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39568 represents the location of a stop codon in AAQ46083.

XX Sequence 1433 AA;

XX Query Match 97.8%; Score 6651; DB 14; Length 1433;

XX Best Local Similarity 97.7%; Pred. No. 0;

Matches 1230; Conservative 7; Mismatches 14; Indels 8; Gaps 2;
QY 1 MELAALCRWGLLALLPPGAASQTCTFNNFTVSEFWLRVPKVSASHLE---LYQCQGVV 56
Db 1 MELAALCRWGLLALLPPGAASQTCTGTDMK---LRLPASPEHDLMLRHLVQVQGVV 56
QY 57 QGNLELYLPTNASISFLQDIOEVQGYVLIAHNQVRQVLPQLRLIRVRGTQLPEDNYALAV 116
Db 57 QGNLELYLPTNASISFLQDIOEVQGYVLIAHNQVRQVLPQLRLIRVRGTQLPEDNYALAV 116
QY 117 LQNGDPLNNTTVPVGTGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHKN 176
Db 117 LQNGDPLNNTTVPVGTGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHKN 176
QY 177 NOLAULTIDTNSRACHPCSPMKSGRCWGESSEDCQSLTRTVACAGGCARCKPLPTDCC 236
Db 177 NOLAULTIDTNSRACHPCSPMKSGRCWGESSEDCQSLTRTVACAGGCARCKPLPTDCC 236
QY 237 HEQCAAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCV 296
Db 237 HEQCAAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCV 296
QY 297 TACPYNLYSTDVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMEHLRVRVAV 356
Db 297 TACPYNLYSTDVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMEHLRVRVAV 356
QY 357 TSANIQEFAGCKKIFGSLAFPLPESPDGPASNTAPLQPELOQVFETLEBITGYLYISAMP 416
Db 357 TSANIQEFAGCKKIFGSLAFPLPESPDGPASNTAPLQPELOQVFETLEBITGYLYISAMP 416
QY 417 DSLPDLVSFQNLQVIRGRILHNGAYSLLTQGLGISWLGRLSRELGSGLALIHNNHLCF 476
Db 417 DSLPDLVSFQNLQVIRGRILHNGAYSLLTQGLGISWLGRLSRELGSGLALIHNNHLSF 476
QY 477 VHTVPWDLFRNPHOALLHTANRDECEVGEGLACHQLCARGHCVGPGTQCVCNCSQFLR 536
Db 477 VHTVPWDLFRNPHOALLHTANRDECEVGEGLACHQLCARGHCVGPGTQCVCNCSQFLR 536
QY 537 GOECVEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPC 596
Db 537 GOECVEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPC 596
QY 597 VARCPGSKPDLSPYMPKPFDEEGACQPCINCTHSCVDLDDKGCAPQASPLTSIYS 656
Db 597 VARCPGSKPDLSPYMPKPFDEEGACQPCINCTHSCVDLDDKGCAPQASPLTSIYS 656
QY 657 AVVGILLVVVLGVVFGILIKRQKIRKYMRRLLQETELVEPLTPSGAMPNQAQMRILK 716
Db 657 AVVGILLVVVLGVVFGILIKRQKIRKYMRRLLQETELVEPLTPSGAMPNQAQMRILK 716
QY 717 ETELKRVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEIIDEAYVMAG 776
Db 717 ETELKRVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEIIDEAYVMAG 776
QY 777 VGSPYVSRLLGICLTSTVQLVTQLMPYGLLDHVRNRLGSLQDILLNWCMIAGMSYL 836
Db 777 VGSPYVSRLLGICLTSTVQLVTQLMPYGLLDHVRNRLGSLQDILLNWCMIAGMSYL 836
QY 837 EDVRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMWALSILR 896
Db 837 EDVRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMWALSILR 896
QY 897 RRTTHOSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPCTIDVYMIW 956
Db 897 RRTTHOSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPCTIDVYMIW 956
QY 957 KCMWIDSECRPRFRELVSERFARMARDPQRFVVIQNEEDLGSPASPLDSTFYRSLEDDDDMGD 1016
Db 957 KCMWIDSECRPRFRELVSERFARMARDPQRFVVIQNEEDLGSPASPLDSTFYRSLEDDDDMGD 1016
QY 1017 LVDABEYLVPQGGFFCPDPAPGAGGVVHHRHSSSTRSGGDLTLGLSESEBAPRSPLA 1076
Db 1017 LVDABEYLVPQGGFFCPDPAPGAGGVVHHRHSSSTRSGGDLTLGLSESEBAPRSPLA 1076

Qy	1077	PSEGAGSDVFDGLGMAAKGLOSLPHDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQ	1136
Db	1077	PSEGAGSDVFDGLGMAAKGLOSLPHDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQ	1136
Qy	1137	PEYVQPDVVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAPGGAVENPEY	1196
Db	1137	PEYVQPDVVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAPGGAVENPEY	1196
Qy	1197	LTPQGGAAPQHPPPPAFSPAFDNLYYWDQDPPPERGAPPTFKGTPTAENPEYLGLDVPV	1255
Db	1197	LTPQGGAAPQHPPPPAFSPAFDNLYYWDQDPPPERGAPPTFKGTPTAENPEYLGLDVPV	1255

Search completed: July 22, 2003, 08:40:12
 Job time : 43.9774 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:25:54 ; Search time 23.0157 Seconds
(without alignments)
5267.077 Million cell updates/sec

Title: SEQ4-59-73-14
Perfect score: 6847
Sequence: 1 MEIAALCRWGILLALLPPGA.....TFKGTPTAENPEYLGLDVPV 1261

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 73.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6707	98.0	1255	1 A24571	protein-tyrosine k
2	5901.5	86.2	1254	2 I48161	p-185 precursor -
3	5901	86.2	1260	1 TVRTNU	protein-tyrosine k
4	3098	45.2	1210	1 GQHUE	epidermal growth f
5	3074	44.9	1210	2 A53183	epidermal growth f
6	3054.5	44.6	1223	1 TVCHLV	epidermal growth f
7	2930.5	42.8	1308	2 A47253	epidermal growth f
8	2633	38.5	1166	1 S06142	protein-tyrosine k
9	2373.5	34.7	1342	2 A36223	kinase-related tra
10	2299.5	33.6	1339	2 JC4387	epidermal growth f
11	1786.5	25.8	698	1 TVFVLV	protein-tyrosine k
12	1703	24.9	604	1 TVYUHF	protein-tyrosine k
13	1650.5	24.1	1330	1 GQFFE	epidermal growth f
14	1647	24.1	544	2 S35745	protein-tyrosine k
15	1640	24.0	545	2 S00727	kinase-related tra
16	1623	23.7	540	2 B4776	protein-tyrosine k
17	1621	23.7	540	1 TVFVFB	protein-tyrosine k
18	1466	21.4	644	2 A36325	epidermal growth f
19	1274	18.6	1323	2 E88257	protein let-23 (im
20	1274	18.6	1374	2 S70712	protein-tyrosine k
21	1186	17.3	1369	2 S70713	protein-tyrosine k
22	1121	16.4	1717	1 A45558	epidermal growth f
23	1086	15.9	527	2 A42032	epidermal growth f
24	942.5	13.8	843	2 A27131	epidermal growth f
25	806.5	11.8	346	2 S13807	protein-tyrosine k
26	754.5	11.0	311	2 S13808	protein-tyrosine k
27	736	10.7	1363	2 T43220	insulin-like growt
28	720	10.5	1382	1 INHUR	insulin receptor p
29	713	10.4	1383	2 A36080	insulin receptor p

RESULT 1

A24571

Protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human

N;Alternate names: c-erb-B-2 protein precursor; kinase-related transforming protein erbB

C;Species: Homo sapiens (man)

C;Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999

C;Accession: A24571; A25491; A44188; B44188; I59509; I57622

R;Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T.;

Nature 319, 230-234, 1986

A;Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth f

A;Reference number: A24571; MUID:86118663; PMID:3003577

A;Accession: A24571

A;Molecule type: mRNA

A;Residues: 1-1255 <YAM>

A;Cross-references: GB:X03363; NID:X031197; PIDN:CAA27060.1; PID:G31198

R;Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.

Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985

A;Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epiderm-

A;Reference number: A25491; MUID:86016729; PMID:2995967

A;Accession: A25491

A;Molecule type: DNA

A;Residues: 737-1031 <SEM>

A;Cross-references: GB:M11767; NID:G182163; PIDN:AAA35808.1; PID:G553282

R;Coussens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg, P.

Science 230, 1132-1139, 1985

A;Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chromos

A;Reference number: A44188; MUID:86070181; PMID:2999974

A;Accession: A44188

A;Molecule type: DNA

A;Residues: 740-910 <COU1>

A;Cross-references: GB:M12036; NID:G183988; PIDN:AAA35978.1; PID:G183989

A;Accession: B44188

A;Molecule type: mRNA

A;Residues: 1-517, 'RALL', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>

A;Cross-references: GB:M11730; NID:G183986

R;King, C.R.; Kraus, M.H.; Aaronsen, S.A.

Science 229, 974-976, 1985

A;Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.

A;Reference number: I59509; MUID:85272597; PMID:2992089

A;Accession: I59509

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 832-909 <REX>

A;Cross-references: GB:L29395; NID:G459807; PIDN:AAA35809.1; PID:G459808

R;Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.

Mol. Cell. Biol. 7, 2597-2601, 1987

A;Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptional

A;Reference number: I57622; MUID:87286898; PMID:3039351

A;Accession: I57622

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-191 <TAL>

A;Cross-references: GB:M16792; NID:g183983; PIDN:AAA58637.1; PID:g553332
 C;Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
 C;Genetics:
 A;Gene: GDB:ERBB2; NGL; NEU; HER-2
 A;Map position: 17q21.1-17q21.1
 A;Introns: 25/1; 75/3; 147/1; 883/3
 A;Note: the list of introns is incomplete
 C;Function:
 A;Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
 A;Superfamily: epidermal growth factor receptor; protein kinase homology
 C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
 inase
 F;1-21/Domain: signal sequence #status predicted <SIG>
 F;22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>
 F;22-653/Domain: extracellular #status predicted <EXT>
 F;70-304/Domain: EGF receptor extracellular domain repeat <EE1>
 F;395-605/Domain: EGF receptor extracellular domain repeat <EE2>
 F;654-675/Domain: transmembrane #status predicted <TM>
 F;676-1255/Domain: intracellular #status predicted <INT>
 F;718-983/Domain: protein kinase homology <KIN>
 F;726-734/Region: protein kinase ATP-binding motif
 F;68.124,187,259,530,571.629/Binding site: carboxydrate (Asn) (covalent) #status predicted
 F;686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
 F;753/Active site: Lys #status predicted
 F;1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 98.0%; Score 6707; DB 1; Length 1255;
 Best Local Similarity 98.3%; Pred. No. 2.9e-267;
 Matches 1240; Conservative 5; Mismatches 10; Indels 6; Gaps 2;

Qy 1 MELAALCRWGLLLALLPPGAASVCTGTDMLRASPETHDMLRHLVGGCVQGNL 60
 Db 1 MELAALCRWGLLLALLPPGAASVCTGTDMLRASPETHDMLRHLVGGCVQGNL 60
 Qy 61 ELTYLPTNASLSFLQDIEQVQ--GVYLAHQVQVPL---QRLIRVGTQLFEDNAL 120
 Db 61 ELTYLPTNASLSFLQDIEQVQ--GVYLAHQVQVPL---QRLIRVGTQLFEDNAL 114
 Qy 121 AVLNDGPDPLNNTPTVTGASPGGLRELQRLSITELKGGVLIQRNPQLCYQDTILWKDIFH 180
 Db 115 AVLNDGPDPLNNTPTVTGASPGGLRELQRLSITELKGGVLIQRNPQLCYQDTILWKDIFH 174
 Qy 181 KNNQALTLIDNRSRACHPCSPCKGSRGWGSEDCQSLTRTVACGGCARCKGPLPTD 240
 Db 175 KNNQALTLIDNRSRACHPCSPCKGSRGWGSEDCQSLTRTVACGGCARCKGPLPTD 234
 Qy 241 CCHQCAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGAS 300
 Db 235 CCHQCAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGAS 294
 Qy 301 CVTACPNYLTVDGSCITLVPLHNQEVTAEDGTQRCBKSKPCARVCYGLGMHLEVR 360
 Db 295 CVTACPNYLTVDGSCITLVPLHNQEVTAEDGTQRCBKSKPCARVCYGLGMHLEVR 354
 Qy 361 AVTSANIQEFAGCKKIFGSLAFLESFDGDPASNTAPLQPEQLQVFLEITGLYISA 420
 Db 355 AVTSANIQEFAGCKKIFGSLAFLESFDGDPASNTAPLQPEQLQVFLEITGLYISA 414
 Qy 421 WPDLSPLDSVFQNLQVIRGRILHNGAYSILTQLGGLISWGLRSLRSLRELGLALIHNTHL 480
 Db 415 WPDLSPLDSVFQNLQVIRGRILHNGAYSILTQLGGLISWGLRSLRSLRELGLALIHNTHL 474
 Qy 481 CFVHTVPWDQLFRPHQALLTANRPEDECVGEGGLACHQLCARGHCWGPGTQCVCNSQF 540
 Db 475 CFVHTVPWDQLFRPHQALLTANRPEDECVGEGGLACHQLCARGHCWGPGTQCVCNSQF 534
 Qy 541 LRQGECEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDP 600
 Db 535 LRQGECEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDP 594
 Qy 601 FCVARCPGSKVDLSYMPINWKFDPDEEGACQPCPINCTHSCVDLDDKGPAPQASPLTSI 660

Db 595 FCVARCPGSKVDLSYMPINWKFDPDEEGACQPCPINCTHSCVDLDDKGPAPQASPLTSI 654
 Qy 661 VSAVVGILLVVLGVVFGIILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAWRI 720
 Db 655 ISAVVGILLVVLGVVFGIILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAWRI 714
 Qy 721 LKETELRKVVLGSGAGFGTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVM 780
 Db 715 LKETELRKVVLGSGAGFGTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVM 774
 Qy 781 AGVSPVSRLLGICLTSTVOLVTQLMPYGLCLDHDVRENRGLGSQLLNWCMQIAKMS 840
 Db 775 AGVSPVSRLLGICLTSTVOLVTQLMPYGLCLDHDVRENRGLGSQLLNWCMQIAKMS 834
 Qy 841 YLEDVRLVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGKGKVPICKMALESI 900
 Db 835 YLEDVRLVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGKGKVPICKMALESI 894
 Qy 901 LRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITDVM 960
 Db 895 LRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITDVM 954
 Qy 961 MVKCMWIDSECRPRFRELSEFSEFMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDM 1020
 Db 955 MVKCMWIDSECRPRFRELSEFSEFMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDM 1014
 Qy 1021 GDLVDABEYLPOQGFPCPDPAAGAGMVHRRSSSTRSGGDLTLGLPSEEEARSP 1080
 Db 1015 GDLVDABEYLPOQGFPCPDPAAGAGMVHRRSSSTRSGGDLTLGLPSEEEARSP 1074
 Qy 1081 LAPSEAGSDVDFDGLGMAAKGLQSLPHTDPSPLOQYSEDPTVPLPSETDGYVAPLTCS 1140
 Db 1075 LAPSEAGSDVDFDGLGMAAKGLQSLPHTDPSPLOQYSEDPTVPLPSETDGYVAPLTCS 1134
 Qy 1141 POPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTSLFGKXGKVVYKDFAFGAVENP 1200
 Db 1135 POPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTSLFGKXGKVVYKDFAFGAVENP 1194
 Qy 1201 EYLTPOGGAPOHPHPAPSPADNLYYWDODPPERCAPPSTKGTGTAEVPEVLGDVP 1260
 Db 1195 EYLTPOGGAPOHPHPAPSPADNLYYWDODPPERCAPPSTKGTGTAEVPEVLGDVP 1254
 Qy 1261 V 1261
 Db 1255 V 1255

RESULT 2
 I48161
 P-185 precursor - golden hamster
 C;Species: Mesocricetus auratus (golden hamster)
 C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999
 C;Accession: I48161
 R;Nakamura, T.; Ushijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishikawa,
 Gene 140, 251-255, 1994
 A;Title: Cloning and activation of the Syrian hamster neu proto-oncogene.
 A;Reference number: I48161; MUID:94193007; PMID:7908275
 A;Accession: I48161
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 1-1254 <RES>
 A;Cross-references: GB:D16295; NID:g493236; PIDN:BA003801.1; PID:g747595
 C;Genetics:
 A;Gene: neu
 C;Superfamily: epidermal growth factor receptor; protein kinase homology
 C;Keywords: ATP
 F;718-983/Domain: protein kinase homology <KIN>
 F;726-734/Region: protein kinase ATP-binding motif

Query Match 86.2%; Score 5901.5; DB 2; Length 1254;
 Best Local Similarity 86.2%; Pred. No. 2.5e-234;
 Matches 1088; Conservative 60; Mismatches 105; Indels 9; Gaps 3;

Qy 1 MELAALCRWGLLLALLPPGAASSTVCTGTDMLKRLPASPTHLDMLRHLHYGCGQVVOGNL 60
Db 1 MELAAWCWGLLLALLSPGASGTCVCTGTDMLKRLPASPTHLDIVRHLHYGCGQVVOGNL 60
Qy 61 ELTYLPTNASLSFLQDIQEVQFNFTVSFMLRVPKVSASHLE-ORLRLVIRGTQFEDNYA 119
Db 61 ELTYLPANATLSFLQDIQEVQ-----GYMLIAHSQVRHVPLQRLRLVIRGTQFEDKYA 113
Qy 120 LAVLDNGDPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCIYQDITLWKDIF 179
Db 114 LAVLDNRDPLNVTATGRTPEGLRELQRLSLTEILKGGVLIQRNPOLCIYQDITLWKDVF 173
Qy 180 HKNQLALTLIDNRSRACHPCSPMKGSRGWSGEDCQSLRTVACGGCARCKGLPT 239
Db 174 RKNQLAPVDITNRSRACPPCAPACKDNHCWGPEDCQTLTGTTIAPRAVPAARLPT 233
Qy 240 DCHEQCAAGCTGPKHSDCLACLPNHSIGICELHCPALVTYNTDTPESMPNPEGRYTFGA 299
Db 234 DCHEQCAAGCTGPKHSDCLACLPNHSIGICELHCPALVTYNTDTPESMPNPEGRYTFGA 293
Qy 300 SCVTACPNYLSLTDVGSCTVLCPLHNOEVTAAEDGTQRCCKSKPCARVCYGLGMEHLREV 359
Db 294 SCVTTCPNYLSLTDVGSCTVLCPLHNOEVTAAEDGTQRCCKSKPCARVCYGLGMEHLRGA 353
Qy 360 RAVTSANIQEPAGCKKIFGSLAFIPESPDGDPASNTAPLOPEQLQVFTETLEITGYLYIS 419
Db 354 RAVTSANIQEPAGCKKIFGSLAFIPESPDGNPSSGIAPLTPPEQLQVFTETLEITGYLYIS 413
Qy 420 AWPDSLPLSVFONLQVIRGILHNGAYSILTOGLGISWLGSLRSLRELSGLALIHNTH 479
Db 414 AWPDSLHDLVSFONLQVIRGILHNGAYSILTOGLGISWLGSLRSLRELSGLVLIHRNTH 473
Qy 480 LCFVHTVPWDQLFRNPQHALLHTANRDECEVCBGSLACHQLCARGHCWGPPTQCVNCSQ 539
Db 474 LCFVHTVPWDQLFRNPQHALLHSGNPSEEEGCLKDPACYPICAGHCWGPPTQCVNCSH 533
Qy 540 FLRGQECVECRVLQGLPREVYNARHCLPCHPECPQNGSVTCFPGPADQVCAHAHYKDP 599
Db 534 FLRGQECVKECRVWKGLPREVYNGKHCLPCHPECPQNGSVTCFPGPADQVCAHAHYKDS 593
Qy 600 PFCVARCPGKVPDLVYMPIKFPDEEGACOPCINCTHSCVDLDDGCPAQBPASPLTS 659
Db 594 PFCVARCPGKVPDLVYMPIKFPDEEGMCPQPCINCTHSCVDLDERGCPAQBPASPLTS 653
Qy 660 IVSAVVGILLVVLGVVFGILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAMR 719
Db 654 IIAIVVGILLFLVIGVVVGLIKRRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAMR 713
Qy 720 ILKETELRKVKVLSGAPGVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYV 779
Db 714 ILKETELRKVKVLSGAPGVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYV 773
Qy 780 MAGVSPVSVLLGICLTSTVOLYTMPLPYCCLLDHYRENRGRIGSQDILLNWCQIAKGM 839
Db 774 MAGLSPVSVRLGICLTSTVOLYTMPLPYCCLLDHYRENRGRIGSQDILLNWCQIAKGM 833
Qy 840 SYLEDVRLVRDLAARNVLKSPNHVKITDFGLARLLDIDTEYHADGKVPKIKWMALES 899
Db 834 SYLEDVRLVRDLAARNVLKSPNHVKITDFGLARLLDIDTEYHADGKVPKIKWMALES 893
Qy 900 ILRRFRTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYM 959
Db 894 ILRRFRTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYM 953
Qy 960 IMVKCMIWIDSCRRPREFLVESEFRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDD 1019
Db 954 IMVKCMIWIDSCRRPREFLVESEFRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDD 1013
Qy 1020 MGDVLDAEYLVLPQOGFPDPAAGAGMVHHRSSSTRSGGGDLTLGLEPSSSEAPRS 1079
Db 1014 MGDVLDAEYLVLPQOGFPDPAAGAGSTARRHRRSSSTRSGGGELTLGMESEEPSPRS 1073
Qy 1080 PLAPSEAGSDVFDGLGMGAALKQLSLTHDPSPLQRYSDPTVPLPSETDGYVAPLTC 1139

Db 1074 PLAPSEAGSDVFEGLGMGATKGPOSISPRDLSPLQRYSEDTPLPTTDTGTVAPLAC 1133
Qy 1140 SPOPEYVQNDVDRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAFGGAVEN 1199
Db 1134 SPOPEYVQNDVDRPQPPSPREGPLPVRPAAGATLERPKTILSPGKNGVVKDVFTFGGAVEN 1193
Qy 1200 PEYLTPOGGAPOHPHPPAFSPAFDNLVYWDQDPPERAGAPPSTFKGTPTTAENPEYLGDV 1259
Db 1194 PEYLVPRGGSASQPH-PPALCPAFDNLVYWDQDPSERGSPPNTFEGTPTTAENPEYLGDV 1252
Qy 1260 PV 1261
Db 1253 PV 1254
RESULT 3
TVRTNU
protein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 11-Jun-1999
C;Accession: A24562; A61204
R;Bargmann, C.I.; Hung, M.C.; Weinberg, R.A.
Nature 319, 226-230, 1986
A;Title: The neu oncogene encodes an epidermal growth factor receptor-related protein.
A;Reference number: A24562; MUID:86118662; PMID:3945311
A;Accession: A24562
A;Molecule type: mRNA
A;Residues: 1-1260 <BAR>
A;Cross-references: EMBL:X03362; NID:956745; PIDN:CAA27059.1; PID:956746
R;Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohen, S.
Carcinogenesis 12, 1975-1978, 1991
A;Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals no m-
2-thiazolyl]formamide or N-methyl-N-nitrosourea
A;Reference number: A61204; MUID:92035293; PMID:1682063
A;Accession: A61204
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 637-663,'V',665-702 <MAS>
A;Note: authors translated the codon GCA for residue 25 as Val
C;Genetics:
A;Gene: neu
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosphor-
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>
F;658-680/Domain: transmembrane #status predicted <TMN>
F;723-988/Domain: protein kinase homology <KIN>
F;731-739/Region: protein kinase ATP-binding motif
F;71,191,263,535,576,634/Binding site: carboxydrate (Asn) (covalent) #status predicted
F;691/Binding site: phosphate (Thr) (covalent) #status predicted
F;758/Active site: lys #status predicted
F;882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted
Query Match 86.2%; Score 5901; DB 1; Length 1260;
Best Local Similarity 86.2%; Pred. No. 2.7e-234;
Matches 1095; Conservative 48; Mismatches 105; Indels 22; Gaps 5;
Qy 1 MELAALCRWGLLLALLPPGAASSTVCTGTDMLKRLPASPTHLDMLRHLHYGCGQVVOGNL 60
Db 4 MELAAWCWGLLLALLSPGASGTCVCTGTDMLKRLPASPTHLDIVRHLHYGCGQVVOGNL 63
Qy 61 ELTYLPTNASLSFLQDIQEVQFNFTVSFMLRVPKVSASHLEORLRLVIRGTQF 113
Db 64 ELTYLPANATLSFLQDIQEVQ-----GYMLIAHSQVRHVPL-----ORLRLVIRGTQF 110
Qy 114 FEDNYALAVLDNGDPLNNTTPVT-GASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDT 172
Db 111 FEDKYALAVLDNRDPQDNVAASPTGRTPEGLRELQRLSLTEILKGGVLIQRNPOLCYQDM 170
Qy 173 ILKWDIFHKNQLALTLIDNRSRACHPCSPMKGSRGWSGEDCQSLTRTVACGGCAR 232
Db 171 VLWKDVFPRKNQLAPVDIDNRSRACPPCAPACKDNHCWGPEDCQTLTGTTCTSGCAR 230

QY 233 CKGLPTDCHEQCAAGCTGPKHSDCLACLPNHSIGICELHCPALVTYNTDTFESMNP 292
DB 231 CKGRLPDCCHEQCAAGCTGPKHSDCLACLPNHSIGICELHCPALVTYNTDTFESMNP 290
QY 293 GRYTFGASCVTACPNYLTVDGSCCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLG 352
DB 291 GRYTFGASCVTTCPNYLTSTVGSCCTLVCPNNQEVTAEDGTQRCCKSKPCARVCYGLG 350
QY 353 MEHLREVRVTSANIQEFAGCKKIFGSLAFIPESFGDPDPAANTAPLQEQLOVETLEEI 412
DB 351 MEHLRGARAITSDNVQEFQCKKIFGSLAFIPESFGDPDPAANTAPLQEQLOVETLEEI 410
QY 413 TGYLYISAWPDSLPDLSVFNQVONLVRGRLHNGAYSLTQGLGILSWLGLRLSRLGSLA 472
DB 411 TGYLYISAWPDSLRDLSVFNQVONLVRGRLHNGAYSLTQGLGILSHLSRLGSLA 470
QY 473 LIHNTHLCFVHTVPWDQFLRNPHQALHTANRPEDE-CVGEGLACHQLCARGHCWGP 531
DB 471 LIHRNAHLCFVHTVPWDQFLRNPHQALHSGNRPEEDLCVSSGLVCNLSCAHGHWCWGP 530
QY 532 TQCVNCSOFLRGQECVCRVLQGLPREYVNAHCLPCHPCQCPQNGSVTCFGEADQCV 591
DB 531 TQCVNCSHFLRGQECVCRVWKGLPREYVSDRKLCPCHPCQCPQNSSETCFGEADQCA 590
QY 592 ACAHYKDPFFCVARCPGSKVDLSYMPIWKFPDEGACQPCPINCTHSCVDLDDKGGPAE 651
DB 591 ACAHYKSSSCVAPCPGSKVDLSYMPIWKFPDEGICQPCPINCTHSCVDLDDKGGPAE 650
QY 652 QRASPLTSIVSAVGVILLVGVVFGILIKRRQKIRKYTMRLLOETELVEPLTPSGA 711
DB 651 QRASPVTFIATVEGVLLFLILVVVGILIKRRQKIRKYTMRLLOETELVEPLTPSGA 710
QY 712 MPNQAMRILKETELRKVKVLSGAPGVYVGIWIPDGENVKIPVAIKVLRNTPSKANK 771
DB 711 MPNQAMRILKETELRKVKVLSGAFGVYVGIWIPDGENVKIPVAIKVLRNTPSKANK 770
QY 772 BILDEAYVMAGVSPYSRLIGICLTSTVOLVTLQMPYGLLDVHRENRGLSGQDLLNW 831
DB 771 BILDEAYVMAGVSPYSRLIGICLTSTVOLVTLQMPYGLLDVHRENRGLSGQDLLNW 830
QY 832 CMQIAGMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVP 891
DB 831 CVQIAGMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVP 890
QY 892 IKWMALESILRRRTHQSDVMSYGVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPP 951
DB 891 IKWMALESILRRRTHQSDVMSYGVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPP 950
QY 952 ICTIDVYMWKWMIDSECPREFELVSEFSRWARDPQRFVWIONEDLGPASPLDSTFY 1011
DB 951 ICTIDVYMWKWMIDSECPREFELVSEFSRWARDPQRFVWIONEDLGPSSPMDSTFY 1010
QY 1012 RSLLEDDMDGLVDAEYLVLPQQGFFCFDPPAPGAGGMVHRRSSSTRSGGDLTLGLEP 1071
DB 1011 RSLLEDDMDGLVDAEYLVLPQQGFFCFDPPAPGAGGMVHRRSSSTRSGGDLTLGLEP 1070
QY 1072 SEEAAPRSLAPSEGAGSDVFDGLGMCAAKGLQSLPHTDPSPIQRYSEDTVPLPSETD 1131
DB 1071 SEEGPPRSLAPSEGAGSDVFDGLGMCAAKGLQSLPHTDPSPIQRYSEDTVPLPSETD 1130
QY 1132 GYVAPLTCSPQEVYNQDVPQPPSPREGPLPAARPAAGATLERAKTLSPCKNGVVKDVF 1191
DB 1131 GYVAPLTCSPQEVYNQDVPQPPSPREGPLPAARPAAGATLERAKTLSPCKNGVVKDVF 1190
QY 1192 AFGGAVENPEYLTQGGAAPOPHPPPAFSDNLYWDQDPPRGPAPPTFKGTPTAEN 1251
DB 1191 AFGGAVENPEYLVREGTASPPHSPAFSDNLYWDQDPPRGPAPPTFKGTPTAEN 1250
QY 1252 PEYGLDVPV 1261
DB 1251 PEYGLDVPV 1260

RESULT 4
GQHU
epidermal growth factor receptor precursor - human
N;Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
C;Species: Homo sapiens (man)
C;Date: 15-Nov-1984 #sequence, revision 27-Nov-1985 #text change 11-Jun-1999
C;Accession: A00641; A25772; S30024; A38672; A00642; A05281; A05281; A60143; A33:
R;Ullrich, A.; Cousens, L.; Hayflick, J.S.; Dull, T.J.; Gray, A.; Tam, A.W.; Lee, J.; Ye,
P.H.
Nature 309, 418-425, 1984
A;Title: Human epidermal growth factor receptor cDNA sequence and aberrant expression of
A;Reference number: A00641; MUID:84219729; PMID:6328312
A;Accession: A00641
A;Molecule type: mRNA
A;Residues: 1-1210 <ULL>
A;Cross-references: EMBL:X00588; NID:g31113; PIDN:CAA25240.1; PID:g757924
A;Note: the authors translated the codon AAG for residue 540 as Asn
R;Ishii, S.; Xu, Y.; Stratton, R.H.; Roe, B.A.; Merlino, G.T.; Pastan, I.
Proc. Natl. Acad. Sci. U.S.A. 82, 4920-4924, 1985
A;Title: Characterization and sequence of the promoter region of the human epidermal growth
A;Reference number: A25772; MUID:85270438; PMID:2991899
A;Accession: A25772
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-29 <ISH>
A;Cross-references: GB:M11234; NID:g181981; PIDN:AAA52370.1; PID:g553272
R;Haley, J.; Whittle, N.; Bennett, P.; Kitchington, D.; Ullrich, A.; Waterfield, M.
Oncogene Res. 1, 375-396, 1987
A;Title: The human EGF receptor gene: structure of the 110 kb locus and identification of
A;Reference number: S30024; MUID:88217333; PMID:3329716
A;Accession: S30024
A;Molecule type: DNA
A;Residues: 1-29 <HA2>
A;Cross-references: EMBL:X06370; NID:g31118; PIDN:CAA29668.1; PID:g31119
R;Haley, J.D.; Waterfield, M.D.
J. Biol. Chem. 266, 1746-1753, 1991
A;Title: Contributory effects of de Novo transcription and premature transcript termination
A;Reference number: A38672; MUID:91107677; PMID:1988448
A;Accession: A38672
A;Molecule type: DNA
A;Residues: 1-29 <HAL>
A;Cross-references: GB:M38425; NID:g181977; PIDN:AAA63171.1; PID:g553271
A;Experimental source: carcinoma cell line A431-7
R;Xu, Y.; Ishii, S.; Clark, A.J.L.; Sullivan, M.; Wilson, R.K.; Ma, D.P.; Roe, B.A.; Merl
Nature 309, 806-810, 1984
A;Title: Human epidermal growth factor receptor cDNA is homologous to a variety of RNAs c
A;Reference number: A00642; MUID:84245835; PMID:6330563
A;Accession: A00642
A;Molecule type: mRNA
A;Residues: 150-187, 'KSUIQAV', 195, 'M', 197, 'A', 199-222, 'S', 224-304, 'RA', 307-321,
, 798-799, 'TD', 802-811, 'R', 813-942 <XUY>
A;Experimental source: A431 human carcinoma cells, which have large numbers of EGF receptor
R;Lin, C.R.; Chen, W.S.; Krueger, W.; Stolarsky, L.S.; Weber, W.; Evans, R.M.; Verma, I.N.
Science 224, 843-848, 1984
A;Title: Expression cloning of human EGF receptor complementary DNA: gene amplification
A;Reference number: A43615; MUID:84196372; PMID:6326261
A;Accession: A43615
A;Molecule type: mRNA
A;Residues: 713-964 <LIN>
A;Experimental source: epidermoid carcinoma cell line A431
R;Simmen, F.A.; Gope, M.L.; Schulz, T.Z.; Wright, D.A.; Carpenter, G.; O'Malley, B.W.
Biochem. Biophys. Res. Commun. 124, 125-132, 1984
A;Reference number: A23062; MUID:85046483; PMID:6093780
A;Accession: A23062
A;Molecule type: mRNA
A;Residues: 1028-1210 <SIM>
R;Weber, W.; Gill, G.N.; Speiss, J.
Science 224, 294-297, 1984
A;Reference number: A05281; MUID:84172183; PMID:6324343
A;Accession: A05281
A;Molecule type: protein
A;Residues: 25-30, 'S', 32-51, 454-467 <WEB>

R; Russo, M.W.; Lukas, T.J.; Cohen, S.; Staros, J.V.
 J. Biol. Chem. 260, 5205-5208, 1985
 A; Title: Identification of residues in the nucleotide binding site of the epidermal growth factor receptor
 A; Reference number: A60143; MUID: 85182650; PMID: 2985580
 A; Accession: A60143
 A; Molecule type: protein
 A; Residues: 740-744, 'X', 746-747 <RUS>
 R; Mroczkowski, B.; Mosig, G.; Cohen, S.
 Nature 309, 270-273, 1984
 A; Title: ATP-stimulated interaction between epidermal growth factor receptor and superoxide dismutase
 A; Reference number: A38023; MUID: 84191554; PMID: 6325948
 A; Contents: annotation; receptor activity
 A; Note: the EGF receptor (and other tyrosine kinases) can nick double-stranded DNA
 R; Chen, W.S.; Lazar, C.S.; Lund, K.A.; Welsh, J.B.; Chang, C.P.; Walton, G.M.; Der, C.J.
 Cell 59, 33-43, 1989
 A; Title: Functional independence of the epidermal growth factor receptor from a domain x
 A; Reference number: A33331; MUID: 90003233; PMID: 2790960
 A; Contents: annotation; internalization signal
 A; Comment: Binding of EGF to the receptor leads to internalization of the EGF-receptor complex
 C; Genetics:
 A; Gene: GDB:EGFR
 A; Cross-references: GDB:120610; OMIM:131550
 A; Map position: 7p12.3-7p12.1
 C; Superfamily: epidermal growth factor receptor; protein kinase homology
 C; Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosphotyrosine
 F; 1-24/Domain: signal sequence #status predicted <SIG>
 F; 25-1210/Product: EGF receptor #status predicted <MAT>
 F; 25-645/Domain: extracellular #status predicted <EXT>
 F; 75-300/Domain: EGF receptor extracellular domain repeat <EE1>
 F; 390-600/Domain: EGF receptor extracellular domain repeat <EE2>
 F; 646-668/Domain: transmembrane #status predicted <TM>
 F; 669-1210/Domain: intracellular #status predicted <INT>
 F; 710-975/Domain: protein kinase homology <KIN>
 F; 718-726/Region: protein kinase ATP-binding motif
 F; 993-1046/Region: coated-pit mediated internalization signal
 F; 1047-1210/Region: inhibitory
 F; 128,175,352,413,444,528,603/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F; 745/Active site: Lys #status experimental

Query Match 45.2%; Score 3098; DB 1; Length 1210;
 Best Local Similarity 48.9%; Pred No. 1,1e-119;
 Matches 622; Conservative 180; Mismatches 356; Indels 114; Gaps 23;

Qy	11	LLALLPPGAA--STQVCTGDMKRLRLPASPETHLMDRLHLYQGVQVQGNLEITYLPTN	68
Db	14	LLAALCPASRALEEKVCQGSNKLTLQGTEDHFLSLQRMFNCEVVLGNLEITYQRN	73
Qy	69	ASLSFLQDIQEVQ--FNNFTVFWLRVPKVSASHLEQRLIRIVRGTLFEDNALAVLNGD	127
Db	74	YDLSFLKTIQVAGVYLIALNTVERIPL-----ENLQIRGNMYENSVALAVLSNYD	126
Qy	128	FLNNTTPVTGASPGSLRELQRLSLTEILKGGVLIQRLNOLPCVQDTILWKDIFHKNNQAL	187
Db	127	-----ANKTLKELPWRNLQELIUGAVRFSNNPNCVSIQWRDVSDFLSNM	176
Qy	188	TLIDNTRSRACHPCSPMCKGRCRGSSESDCCSLTRTYCAGGCA-RCKGPIPTDCCHQC	246
Db	177	SMDFQNLHSGCQKCDPPSCNGSCWAGENGCKLTKIICAQCCGRCKGSPSCCHNQ	236
Qy	247	AAGCTGPKHSDCLACLFHNSGI CELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP	306
Db	237	AAGCTGPRSDCLVCRKFRDEATCKDTCPLMLYNPTTYQMDVNPDEGKYSFGATCVKKCP	296
Qy	307	YNILSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARCYGLGHEHLREVAVTASN	366
Db	297	RNYVTDHSGCVACGADSYEM-EDGGRKCKCKGECGRKVCNGIGIGFEKDSUSINATN	355
Qy	367	IQEFAGCKKIFGSLAFPESDGDPASNTAPLQEQVFTLEBITGYLYISAWPDSLP	426
Db	356	IKHFNKNTSISGDLHILPVARFGDSFTHTPPLDQELDLKTVKEITGLFLIQWPENRT	415
Qy	427	DLVSFQNLQVIRGRILLINGAYSLTLQGLIGISWLGRLSRLRELGSGLALHHNTHLCFVHT	486

RESULT 5

A53183

epidermal growth factor receptor precursor - mouse
 C; Species: Mus musculus (house mouse)

C; Date: 06-Jan-1995 #sequence revision 06-Jan-1995 #text change 18-Jun-1999

C; Accession: A53183; A43818; S24942; A28941; S45325; I49643

R; Luetteke, N.C.; Phillips, H.K.; Qiu, T.H.; Copeland, N.G.; Earp, H.S.; Jenkins, N.A.;

Genes Dev. 8, 399-413, 1994

A; Title: The mouse waved-2 phenotype results from a point mutation in the EGF receptor

A; Reference number: A53183; MUID: 94170986; PMID: 8125255

A; Accession: A53183

A; Molecule type: mRNA

A; Residues: 1-1210 <LUE>

A; Cross-references: GB:U03425

Db	416	DLHAFENLEIRGRTKQHQGFSLAVSLNTITSGLRSLKLSIDGDIVISGNKNLCYANTI	475
Qy	487	PWDLFRNFHQALLHTANRDECEVGLACHOLCARGHCWGPGPTOCVNCQFRLGOEC	546
Db	476	NWKLFGTSGQTKIISNRGENSKATGOVCHALCSPGCGWGPDRDCVSCRNVSRGEC	535
Qy	547	VEBCRVLOGLPREYVNAHCLPCHPEQOPQNGSVTCFPGPADQCACAHYKDPFPFCVARC	606
Db	536	VDKCKLGEPRFEVENSEICQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTC	595
Qy	607	PSGVKPLDSYMPINWPKFDEBEGACOPCPINCHSCVLDLDDKCCPAEQRASPLTSIVSAVG	666
Db	596	PAGVGGNNNTL-VNKYADAGVHCHLCPNCTYCTGPLEGCPNTPGPKIP--STATGAVG	652
Qy	667	---ILLVVVLGVVFGIILKRRQKIRYTMRRLLQETELVEPLTPSGAMPNQAOIRLKE	723
Db	653	ALLULLLVVAGIG---LFMERRRHVVKRTLRLQLRELVEPLTPSGEAPNQLLRILKE	709
Qy	724	TELRKVKVLGSGAPGVYKIGIWDGNGVKIPVAIKVLRNTSPKANKEILDEAYWAGV	783
Db	710	TEPKKIKVLGSGAGFTVYKGLWIPGEGKVKIPVAIKELREATSPKANKEILDEAYWASV	769
Qy	784	GSPVVSLLGLCTSTVOLTVMPLPYCILDHYRNRGLGSQDLNMCWQIAKMSVLE	843
Db	770	DNPHVCELLGLCTSTVQLITQLMPPFCGLDYYREHKDNTGSOYLLNMCVQIAKGMYLE	829
Qy	844	DVRLVHRDLAARNVLKSPNHVKITDFGLARLIDIDETEHADGGKVPKIMWALESLRR	903
Db	830	DRRLVHRDLAARNVLVKTPOHVKITDFGLAKLGAEEKEYHAEKGKVPKIMWALESLHR	889
Qy	904	RFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVK	963
Db	890	IYTHQSDVWSYGVTVWELMTFGSKPYDGIIPASSISSILEKGERLPQPPICTIDVYIMVK	949
Qy	964	CWIDSECRPRRELVSFSEFMRMDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDDMDGD	1022
Db	950	CWMDADSRPKRELIIIEFSKMDRDPQRYLVIQGDREHMLPSTDSNFYALMDEEDMD	1009
Qy	1023	LVDAAEYLVFPQGGFCDDPAPGAGGMVHRHRSSTRSGGDLTLGLEPESEEAAPRPLA	1082
Db	1010	VVDADEVLIIPQGGFF-----SSPSTSRPTLL	1035
Qy	1083	PSEGAGSDVDFDGLMGAAKGLQSLPDPSPLOQYSEEDPTVPLPSET--DGVVAPLTCS	1140
Db	1036	SSLATSNN--NSTVACIDRNLQSCPIKEDSFLOQYSSDPTGALTEDSIDDTFL-----	1087
Qy	1141	PQBEYVNPQDVRQPPSPREGPLPAAPAGATLERAKTLPKNGVGVKVPFAGGAVENP	1200
Db	1088	PVPEYINQ-SVPKRPAGSVQNPVYHNQPLNP-----APSRDPHYQD--PHSTAVGNP	1136
Qy	1201	EYL-TPGGAAAPQHPHPPAFSPAFDNLVYWDQ-----DP-----PERGAPPSTF	1243
Db	1137	EYLVNTVO-----PTCVNSTFDSNAHWAQKSHQISLDNPDYQDQDFPKEAKPNGIF	1187
Qy	1244	KGTPTAENPEYL	1255
Db	1188	KGS-TAENAEYL	1198


```

RESULT 6
TVCHLV
epidermal growth factor receptor precursor - chicken
N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
C:Species: Gallus gallus (Chicken)
C:Date: 28-Feb-1986 #sequence_revision 05-May-1995 #text_change 04-Feb-2000
C:Accession: A27720; A00643
R:Lax, I.; Johnson, A.; Howk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Vennstr
Mol. Cell. Biol. 8, 1970-1978, 1988
A:Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mou
A:Reference number: A27720; MUID:82261272; PMID:3260329
A:Accession: A27720
A:Molecule type: mRNA
A:Residues: 1-1223 <LAX>
A:Cross-references: GB:M20386
R:Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M
Cell 41, 719-726, 1985
A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A:Reference number: A00643; MUID:85228222; PMID:2988784
A:Accession: A00643
A:Molecule type: mRNA
A:Residues: 585-1223 <NIL>
A:Cross-references: GB:M10066
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
specific protein kinase
F:1-30/Domain: signal sequence #status predicted <SIG>
F:31-1223/Product: epidermal growth factor receptor #status predicted <MAT>
F:31-654/Domain: extracellular #status predicted <EXT>
F:81-307/Domain: EGF receptor extracellular domain repeat <EE1>
F:397-610/Domain: EGF receptor extracellular domain repeat <EE2>
F:655-677/Domain: transmembrane #status predicted <TM>
F:678-1223/Domain: intracellular #status predicted <INT>
F:719-984/Domain: protein kinase homology <KIN>
F:727-735/Region: protein kinase ATP-binding motif
F:136,202,280,361,370,422,570,580,615,635/Binding site: carboxydrate (Thr) (covalent) #s
F:192,650/Binding site: carboxydrate (Ser) (covalent) #status predicted
F:687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:754/Active site: Lys #status predicted
F:1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat
Query Match 44.6%; Score 3054.5; DB 1; Length 1223;
Best Local Similarity 48.0%; Pred. No. 6.6e-118;
Matches 625; Conservative 177; Mismatches 350; Indels 151; Gaps 28;
Qy 8 RWGILLALLPPGAA-----STQVCTGDMKRLRPASPETHLMDLRHLYQGCQVQGNLE 61
Db 13 RGAAVLVLLGLGVALCSAVEKKVCQGTNNKLTQLGHVEDHFTSLQRMNNCEVVLNLE 72
Qy 62 LTYLPTNASTLQDIQSVQNNFTVSPWLRVPKVSASHLEQRLRIVRGTOLPEDNYALA 121
Db 73 ITYEHRNRLTLTKTIQEVA--GVVL---IALNMVDVIPLE-NIQIRGNVLYDNSFALA 126
Qy 122 VLDNGDPLNNTPVTGASPGGLRELQSLRLTEILKGLVLIQRIQNPOLCYQDTILWKDIFHK 181
Db 127 VLSNVH-MNKQT-----GLRELPMKRLSEILNGGVKISNNPKLCNMOTVLNNDIIDT 177
Qy 182 NNQALTLID-TNRSRACHPCSPCKGSRCKWGESSEDCQSLTRTVACAGCA-RCKGPLPT 239
Db 178 SRK-PLTVLDFASNLSSCPKHPNCTEDHCWGAGEQNCQTLTKVICAQCCSGRCRKVPS 236
Qy 240 DCCHQCAAGCTGPKHSDCLALHNHSGICELHCPALVTYNTDTPFSMPNPEGRYFGA 299
Db 237 DCCNNQCAAGCTGPRSDCLACRFDRDATCKDTCPPLVLYNPTYQMDVNPESKYSFGA 296
Qy 300 SCVTACPNYLSLTDVSGCTLLCPHLNHOEVTAEADGTORCEKSKPCARVCYGLGMEHLREV 359
Db 297 TCVRECPHNVVTDHSGVRSNCNTDTYEV-BENGVRKCKKCDGLGCKVNGIGIGELKGI 355
Qy 360 RAVTSANIQEPAGCKIFGSLAFIPESFGDGPASNTAPLQPEQLQVFETLEITGYLIYS 419

```

```

Db 356 LSINATNIDSFKNCTKINGDVSLPVAFLGDAFTKTLPLDPKLDVFTVKESGFLLIQ 415
Qy 420 AWPDSLPLSVFONQVIRGRILHNGAYSLLTQGLIGISWLGSLRSLRELGLSLALIHNTH 479
Db 416 AWPDNATDLVAFENLEIRGTQKHGOYSIAVNLKIOSLGLSLKESISDGDIAIMKNKN 475
Qy 480 LCFVHTVPMDQLFRNPQHALLHTANPEDECVGEGGLACHOLCARGHCWGPGPTQCVNCSQ 539
Db 476 LCYADTNWME-SLFATQSQTKIIQNRNKNDCADRHRVCDPLCSDVGCWGPFPFCRCRF 535
Qy 540 FLQGECEVSCRVLQGLPREYVNAHCLPCHPECPQNG---SVTCRGPADQCVACAHY 596
Db 536 FSRQKECVKCNILQGEPRERFDSKCLPCHSECLVQNSTAYNTTCSGPGPDHCKMCAHF 595
Qy 597 KDPFFCVARCPGSKVPDLSYMPYWKPFDEGACQPCPINCTHSCVDLDDGCGPAEQRASP 656
Db 596 IDGPHCVKAPAGVLGENDTL-VKXADANAVCOLCHPNCTRGCKGFLGECGP---NGSK 651
Qy 657 LTSIVSAVV-GILLVVVLGVVFGILIKRRQKIRKYTMRLLOETELVEPLTPSGAMPNQ 715
Db 652 TPSIAAGVGGLLCLLVVGLGIGLYLRRR-HIVRKRTLRLLOERELVEPLTPSGEAPNQ 710
Qy 716 AQMRILKETELRKVKVLGSGARFTVYKGIWIPGENVKIPVAIKVLRENTSPKANKILD 775
Db 711 AHLRIKETEFKKVKVGLSGAFGTVYKGLWIPGEKVKIPVAIKELREATSPKANKILD 770
Qy 776 EAYMAGVGSPPYSLRILGICLTSTVOLATOLMPYVGLLDHVNRNRLGSGDILLNWCMI 835
Db 771 EAYMASVDNPHVCRLLGICLTSTVQITOLMPYVGLLDYIREHKONIGSOYLLNWCMI 830
Qy 836 AKGMSYLEDVRLVHRDLAARNVLVKSNNHVKITDFGLARLLDIDETEHADGGKVPKWM 895
Db 831 AKGMNLEERLVRDLAARNVLVKTPOHVKITDFGLAKLGADEKEYHAEGGKVPKWM 890
Qy 896 ALSEILARRTHOSDVWSYGVTVWELMTFGAKYDGIPIAREIPDLLEKGBRLPOPPCTI 955
Db 891 ALSEILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIPIASEISSVLEKGBRLPOPPCTI 950
Qy 956 DVTMIMVKMWIDSECRPRELVSFRSMARDPQRVVIQ-NEDLGPASPLDSTFVRS 1014
Db 951 DVTMIMVKMWIDADSRRKRELIAEFSKMARPPRYLVIOGDERMHLPSPTDSKVFRTL 1010
Qy 1015 LEDDDMGDLVDABEYLVPOQFFCPDPAPGAGGMVHRHRSSTRSGGDLTLGLEPSEE 1074
Db 1011 MEEDMEDIVDAEYLVPHQGF-----NSPST----- 1038
Qy 1075 EAPRSL-----APSEGAGSDVFDGLGMAAGLQSLPHDPSPLORYSEDPVLPSE 1129
Db 1039 --SRTPLLSLSATSNNATNCID-----RNGOCHPVREDSFVORYSSDPTGNFLEE 1088
Qy 1130 T--DGVAVPLTCSPOEYVNPQDVRPOPSPRESGLPAARPAGATLERAKTLPKNGV 1187
Db 1089 SIDGEL-----PAPEYVNO--LMPKKPS-----TAMVQOIY 1119
Qy 1188 KDVF-----AFGAVENPEYLTPOGGAAPOPHPPPAFSPAFNDLYWDQ- 1231
Db 1120 NNISLTAISKLPMDSRVQNSHSTAVDNPEYL-----NTNOSPALKTVFESSPYIQS 1171
Qy 1232 -----DPPE-----RGAPSTFKGTPTAENPEYLGLDVP 1260
Db 1172 GNHQINLDNDPYQDDFLPNETKPNGLLKVPAAENPEYLRVAAP 1214

```

RESULT 7

A47253

epidermal growth factor receptor, HER4 - human

C:Species: Homo sapiens (man)

C:Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999

C:Accession: A47253

R:Florman, G.D.; Culouscou, J.M.; Whitney, G.S.; Green, J.M.; Foy, L.; Ne

Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993

A:Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epidermal

Db 226 CCHDECAGCGSPQDTCDFACRHFENDSGACVPRCPQPLVYNKLTFOLEPNPHTKYQYGV 285
Qy 301 CVTACPNYSLTDVSGCTLVCLPHNQYTAEDGTQRCCKSPCARCYGLGMEHLREVR 360
Db 286 CVASCPHFNV-VDQTSVRCAPPPDKMEVD-KNGLKMCPCGGLCPKACEGTGSG--SRFQ 341
Qy 361 AVTSANIOFAGCKKIFGSLAFIPESFDGDPASNTAPLOQLOVFTLEBITGYLYISA 420
Db 342 TVDSNIDGFVNCYKILGNLDFLTGLNGDPWHKIPALDPEKLVFRVIRITGYLIQOS 401
Qy 421 WPDLSPLDSVFONQVIRGRILHNGAYS-LTQGLGISWGLRLSLRELGLALIHNNHT 479
Db 402 WPPHMHNFVSFNLTTIGRSLYNRGFSLLIMKNLNTSLGFRSLKETSAGRIVISANRQ 461
Qy 480 LCFVHTVPWDQLFRNPHQALLHTA-NRPEDECVGEGLACHOLCARGHCWGPGTQCNCVS 538
Db 462 LCYHSLNMTKVLRGPTPEERLDIKHNRPRDCVAEGKVCPLCSGGCGWPGPQCLSCR 521
Qy 539 QFLRGOECVECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFPGPADOCVACAHYKD 598
Db 522 NYSRGVCVTHCNFLNGEPREFAHEACFSCHPECPQMEGTATCNGSGSDTCAQCAHFRD 581
Qy 599 PPFVCARCPGVKPDLSYMPIWKPDPDEGACQPCPINCHSCVDLDDKGCAPAEQRA---- 654
Db 582 GPHCVSSCPHGVLG--AKGPYKYPDVQNECRPCHENCCTQCKGPELQDCLGLTLVLIGK 639
Qy 655 SPLTSIVSAVGIILVVLGVVFGILLIKRQOKIR-KYTMRLLOETELVELPELTPSGAMP 713
Db 640 THLTWALTVIAG--LVVIFMLGGTFLYWRGRRIONKRAMRYLERGESIEPLDPS-EKA 696
Qy 714 NOAQRILKTELKRVKVLGSGAFVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEI 773
Db 697 NKVLARIKTELRLKVLGSGVFTGKGVNIPGESIKIPVCKLVIEDSKGSGQFQAV 756
Qy 774 LDEAYMAGVSPYVSRLLGICLTSTVOLVQLMPYGCGLLDHVRNRRGLSGDQLNWCW 833
Db 757 TDHMLAIGSLDHAHVLRLGLCPGSSQLVLTQYPLGSLLDHVRHGRGALPQLLNWGV 816
Qy 834 QIAKMSVLEDRVLRHRLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVIK 893
Db 817 QIAKMYLEBHGWMHRLAARNVLKSPSQVQVADFGVADLLPDDDKOLLYSEAKTPIK 876
Qy 894 WNALESILRRFTHOSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPIC 953
Db 877 WNALESIFHKYTHOSDVWSYGVTVWELMTFGAEPYAGRLAEVFPDLLEKGERLAQOIC 936
Qy 954 TIDVTMVMKCMIDSECRPRELVSFSEFMRARDPQRFVVIQNEDLGPA---SPLDSTF 1010
Db 937 TIDVTMVMKCMIDENIRPTEKELANEFTRMARDPPRYLVIKRES-GPGIAPGPEPHGL 995
Qy 1011 YRSLLDDMDGLVDAAEYLVPOQGFCCPDPAAGAGVHHRSSSTRSGGDLTLGLE 1070
Db 996 TNKKLEEVLEPELDLDLEABED-----NLATTIGLSALSPLVG 1036
Qy 1071 P-SEEAAPRSLAPSEGAGSVDFDGLGMGAAGLQSLPTH-D-PSPLQRYSEDPTVLP- 1127
Db 1037 TLNRRPQSQSLSPSGY-MPMNQNLGESCQESAVSGSSRCRPPVSLH-----PMR 1089
Qy 1128 -----SETDGYVA-----PLTCSQPE-----YVNPQDVRPQPPSPREGP 1162
Db 1090 GCLASESGHVTGSEAELEQKVMCRSRSRSPRPRGDSAYHSQRHSLLTPVTLSP 1149
Qy 1163 -----LPAARPAGATLERAKTLP-SKNGV-----KDVFAFGAVENPEYLT 1204
Db 1150 GLEEDVNGVYMPDTHLKTGTPSSREGTLSSVGLSSVLGTBEDEDE-----BEYEYMN 1201
Qy 1205 POGGAAPOPHPPPAPSPAFDNLVYWD-----QDPPERGAPSPTEKGTPTAENP 1252
Db 1202 RRRHSP-PHPRPSSLSSELGYEYMDVGSDSLASLGSTQSCPLHPVPIMPAGTTPDEY 1260
Qy 1253 EYL 1255
Db 1261 PYM 1263

RESULT 10

JC4387
epidermal growth factor receptor homolog precursor - rat
N;Alternate names: ErbB3 protein; HER3 protein
C;Species: Rattus norvegicus (Norway rat)
C;Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998
C;Accession: JC4387
R;Hellyer, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.
Gene 165, 279-284, 1995
A;Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protein.
A;Reference number: JC4387; MUID:96096535; PMID:8522190
A;Accession: JC4387
A;Molecule type: mRNA
A;Residues: 1-1339 <HEL>
A;Cross-references: GB:U29339; NID:G915389; PID:G915390
A;Experimental source: liver
A;Note: The authors translated the codon AAC for residue 369 as Thr and GTT for residue -
C;Comment: This protein is a functional heregulin receptor that transduces signals to the
C;Genetics:
A;Gene: ErbB3
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homologs;
C;Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane protein
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1339/Product: epidermal growth factor homolog #status predicted <MAT>
F;640-659/Domain: transmembrane #status predicted <TM>
F;705-970/Domain: protein kinase homology <KIN>
F;713-721/Region: protein kinase ATP-binding motif
F;939,1051,1156,1194,1196,1219,1257,1259,1273,1286,1325/Binding site: phosphate (Tyr) (cc
Query Match 33.6%; Score 2299.5; DB 2; Length 1339;
Best Local Similarity 40.2%; Pred. No. 5.4e-87;
Matches 521; Conservative 170; Mismatches 431; Indels 173; Gaps 37;
Qy 3 LAALCRWGLLLALLPPGAA---STQVCTGTDMKRLRPASPETHLDMLRHLYQGCQVVOGN 59
Db 7 LQVLQ----FLLSLARGSEMGNSQAVCPGLNGLSVGTGDADNQVQTLKLYEKEVVMGN 62
Qy 60 LELTYLTNLSLFLQDIQIE-----VQFNFTVSFWRVPKVSASHLEORLRIVRTQL 113
Db 63 LEIVLTGHNADLSFLOWIREVTAYVLVAMNEFSV---LPLP-----NLRVVRGTQV 110
Qy 114 FEDNYALAVLDNGDPLNNTTPVTGASPGGLREQLRLSLTEILKGGVLIORNPOLCYODTI 173
Db 111 YDKGFALFV-----LWNT-----NSSHARLQKLTQLTTEILSGGVIEKNDKLCHWDTI 161
Qy 174 LWKDIHFKNQALTLTIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRITVCAGGC-AR 232
Db 162 DWRDIVRVR--GAELVVKNGANGCPPCHEVCKG-RCWGPDPDQLTKTICAPQCNGR 217
Qy 233 CKGPLPTDCCHEQCAAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDTFESMENPE 292
Db 218 CFGPNPNQCCHEDECAGCGSPQDTCDFACRHFENDSGACVPRCPQPLVYNKLTFOLEPNPH 277
Qy 293 GRVTFGASCVTACPNYSLTDVSGCTLVCLPHNQYTAEDGTQRCCKSPCARCYGLG 352
Db 278 TKYQYGVGVASCPHFNV-VDQTSVRCAPPPDKMEVD-KHGLKMCPCGGLCPKACEGTG 335
Qy 353 MEHLREVRVATSANIOFAGCKKIFGSLAFIPESFDGDPASNTAPLOQLOVFTLEEI 412
Db 336 SG--SRVQTVDSNIDGFVNCYKILGNLDFLTGLNVDPMHKIPALDPEKLVFRVIREI 393
Qy 413 TGILYISAWPDSLPLDSVFONQVIRGRILHNGAYS-LTQGLGISWGLRLSLRELGSGL 471
Db 394 TGYLNIQSWPPHMHNFVSFNLTTIGRSLYNRGFSLLIMKNLNTSLGFRSLKETSAGRI 453
Qy 472 ALTHNNTHLCFVHTVPWDQLFRNPHQALLHTA-NRPEDECVGEGLACHOLCARGHCWPG 530
Db 454 VYISANQOLCYHSLNMTKVLRGPTPEERLDIKYDRPLGECLAEKVCPLCSGGCGWPA 513
Qy 531 PTCVNCQSQFLRGOECVECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFPGPADOC 590

Db 514 PQQCLSCRYNREGVCTVCHNPLQGEPRFVHEAQCFCCHPECLPMEGTSTYNSGGSDAC 573
Qy 591 VACAHYKPPFCVACPCPGVKPDLSPYIMKFPDDEGACQCPINCTHSC--VLDLDDKGC 648
Db 574 ARCAHFRDGPCHVNSCPHILG--AKGPIYKVPDAQNECRPCHENCITQCGNGPELQDCLG 631
Qy 649 PAEQBASPLTISVANVGLILLVVLGVVFGVILIKRQOKIR-KYTMRLLOETELVEPLT 707
Db 632 QAEVLMKPHIVIAVTVG--LAVILMLGGSFLYWRGRIIONKRAMRYLERGESIEPLD 689
Qy 708 PSGAMPNOAQRILKETELRKVKVLGSGAFGVYKGIWIPGENVKIPVAIKVLRENTSP 767
Db 690 PS-EKANKVLARIKETELRKVKVLGSGVFGTVHKGIWIPGESIKIPVCIKVIEDKSGR 748
Qy 768 KANKEILDEAYVMAGVSPYVSRLLGICLTSTVOLVQTMPLPYGCLLDHVRENRLGSDQ 827
Db 749 QSFOAVTDHMLAVGSLDAHVRLLLGLCPGSSQLVQVLYPLGLSLLDHVQKHRETGLPQL 808
Qy 828 LLNWCQIAKMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADG 887
Db 809 LLNMGVQIAKMGYILEEHSVMHRDLAENVMKSPSQVQVADFGVADLLPDDKQLLHSE 868
Qy 888 GKVPDKMALESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERL 947
Db 869 AKTPIKMALESIIHFCKTHQSDVMSYGVTVWELMTFGAEFYAGLRLAEIPDLLEKGERL 928
Qy 948 PQPPICTIDVYMWKMWIDSECRPRRELVSFSEFMRARDPQRFVVTIONEDLGPASPLD 1007
Db 929 AQPQICTIDVYMWKMWIDENIRPTPKELANEFTRWARDPPRYLVIKRAS-GPGTP-- 985
Qy 1008 STFRSLLEDMDGLVDAAEYLVPOQGFPCDPAPGAGGVVHHRHRSSTRSRGGDLTL 1067
Db 996 PAAPFSVLTTKEL--QEALEPEL-----DLDL 1011
Qy 1068 GLEPSEE-----EAPSPAPSEG-----AGSDVFGDLGMC 1099
Db 1012 DLEAEGLATLSGALSALPTGTLTRPGSOLSLSPSSGYPMQSSLGEACLDSAVLGG 1071
Qy 1100 AAKGLQSLTHDPSPLOQRYSEDPTVLPSETDGVY----APL-----TC-----SPQ 1143
Db 1072 REQFSRPISLH-PIPRGR-----PASESEGHVTGSEAELOEKVSVCRSRSRSPRP 1123
Qy 1144 E-----YVNPQVVRPQSPSPREGP-----LPAARPAGATLERAKTLSP-GKNGV 1187
Db 1124 RGSAYHSQRHSLTPTVPLSPPGLEEDNGYVMPDTHLRGASSRSRGTLLSVGLSSVL 1183
Qy 1188 -----KDVFAFGAVENPEYLTPOGGAPOPHPP 1216
Db 1184 GTEEDED-----EEYEMNRKRKRGSP-PRPP 1209

RESULT 11
TVFVLV
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian leukosis virus
N:Contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase
C:Species: avian leukosis virus, ALV
C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 11-Jun-1999
C:Accession: B00643; A00643
R:Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1985
A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A:Reference number: A00643; MUID:85228222; PMID:2988784
A:Accession: B00643
A:Molecule type: mRNA
A:Residues: 1-698 <NLI>
A:Cross-references: GB:M10066; GB:M13881; NID:g211749; PIDN:AAA48763.1; PID:g211750
A:Note: In Genbank entry CHKERBBF, release 109.0, the source is designated as Gallus gal
C:Comment: This protein is synthesized as a gag-env-erbB protein.
C:Genetics:
A:Gene: gag-env-erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific P
F:1-6/Product: gag protein (fragment) #status predicted <GAG>

F:7-59/Product: env protein (fragment) #status predicted <ENV>
F:60-698/Product: protein-tyrosine kinase erbB #status predicted <ERB>
F:194-459/Domain: protein kinase homology <KIN>
F:202-210/Region: protein kinase ATP-binding motif
F:229/Active site: Lys #status predicted

Query Match 25.8%; Score 1766.5; DB 1; Length 698;
Best Local Similarity 52.2%; Pred. No. 1.8e-65;
Matches 374; Conservative 80; Mismatches 137; Indels 125; Gaps 18;

Qy 584 GPEADQCVACAHYKDPFCVACPCPGVKPDLSPYIMKFPDDEGACQCPINCTHSCVDL 643
Db 60 GP--DHCKCAHFIDGPHCVKACPAVLGENDTL-VMKYADANAVCOLCHPNCNTRCGKP 116
Qy 644 DDKCPAEQAEASPLTISVSAV-GILLVVLGVVFGVILIKRQOKIRKYTMRLLOETEL 702
Db 117 GLECP---NGSKTPSTAAGVVGGLLCLLVVGLGIGLYLRR-HIVKRTLRLLOEREL 172
Qy 703 VEPLTPSGAVPNOAQRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVL 762
Db 173 VEPLTPSGEAPNOAHLRIKETEPFKVKVLGSGAFGVYKGLWIPGEKVKIPVAIKEL 232
Qy 763 ENTSPKANKEILDEAYVMAGVSPYVSRLLGICLTSTVOLVQTMPLPYGCLLDHVREN 822
Db 233 EATSPKANKEILDEAYVMAGVSPYVSRLLGICLTSTVOLVQTMPLPYGCLLDVIREH 292
Qy 823 LGSODLLNWCQIAKMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETE 882
Db 293 IGSQYLLNWCQIAKMSYLEDVRLVHRDLAARNVLKTPQHVKITDFGLAKLGADEKE 352
Qy 883 YHADGGKVPKIKMALESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDL 942
Db 353 YHAEKGKVPKIKMALESILRRRTHQSDVMSYGVTVWELMTFGSKPYDGIIPAREIPDL 412
Qy 943 KGERLPQPPICTIDVYMWKMWIDSECRPRRELVSFSEFMRARDPQRFVVIQ-NEDLG 1001
Db 413 KGERLPQPPICTIDVYMWKMWIDADSRPKRELIAEFSKWARDPPRYLVIOGDERMH 472
Qy 1002 PASPLDSTFRSLLEDMDGLVDAAEYLVPOQGFPCDPAPGAGGVVHHRHRSSTRSR 1061
Db 473 LPSTDSKFRVTLMEEDMEDIVDAEYLVPHOGFF-----NSPST--- 513
Qy 1062 GGDLTLGLEPSEBEAPRSL-----APSEAGSDSDVFGDGLGMAAKGLQSLTHDPSP 1116
Db 514 -----SRTPLLSLSATSNNSSATNCID-----RNGQGHVPVREDSFVQ 550
Qy 1117 RYSEDPTVLPSEET--DGYVAPLTCSPQPEYVNPQDVRPQPPSPREGPLPAARAGATLE 1174
Db 551 RYSSDPTGNFLSESIDDGLF-----PAPEYVNO--LMPKKPS----- 585
Qy 1175 RAKTLPSPKXGVVKDVF-----AFGGAVENPEYLTPOGGAAPQPHPPAF 1219
Db 586 -----TAMVQNIYNNISLTAISKLPMSRQNSHSTAVDNPEYL-----NTNQSLA 633
Qy 1220 SPADNLYYWDQ-----DPPE-----RGAPPSTFGKTPTAENPEYLGLDVP 1260
Db 634 KTVFESSPYWQSGNHQINLNDPDYQDQDFLPNETKPNGLLKVPAAENPEYLRVAA 689

RESULT 12

TVYUH

protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain H)
C:Species: avian erythroblastosis virus
C:Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 11-Jun-1999
C:Accession: A00644; A38022
R:Yamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.
Cell 35, 71-78, 1983
A:Title: The erbB gene of avian erythroblastosis virus is a member of the src gene family.
A:Reference number: A00644; MUID:84026539; PMID:6313229
A:Accession: A00644
A:Molecule type: DNA
A:Residues: 1-604 <YAM>
A:Cross-references: GB:K01216; NID:g209676; PIDN:AAA42400.1; PID:g209678

R;Debuire, B.; Henry, C.; Benaissa, M.; Biserte, G.; Claverie, J.M.; Saule, S.; Martin, Science 224, 1456-1459, 1984
A;Title: Sequencing the erpA gene of avian erythroblastosis virus reveals a new type of A;Reference number: A38022; MUID:84223957; PMID:6328658
A;Accession: A38022
A;Molecule type: DNA
A;Residues: 1-28, 'W', 30-139, 'F', 141-145, 'V', 147-152 <DEB>
A;Cross-references: GB:K02006
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific P
F;130-395/Domain: protein kinase homology <KIN>
F;138-146/Region: protein kinase ATP-binding motif
F;165/Active site: Lys #status predicted

C;Species: Drosophila melanogaster
C;Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 11-Jun-1999
C;Accession: A00640; A38021
R;Livneh, E.; Glaser, L.; Segal, D.; Schlessinger, J.; Shilo, B.Z.
Cell 40, 599-607, 1985
A;Title: The Drosophila EGF receptor gene homolog: conservation of both hormone binding
A;Reference number: A00640; MUID:85124611; PMID:2982499
A;Accession: A00640
A;Molecule type: DNA
A;Residues: 1-1330 <LIV>
A;Cross-references: EMBL:K03054
R;Wadsworth, S.C.; Vincent III, W.S.; Bilodeau-Wentworth, D.
Nature 314, 178-180, 1985
A;Title: A Drosophila genomic sequence with homology to human epidermal growth factor re
A;Reference number: A38021; MUID:85137938; PMID:2983232
A;Accession: A38021
A;Molecule type: DNA
A;Residues: 'A', 832-866, 'V', 868-943, 'QTPSLVK' <WAD>
A;Cross-references: EMBL:X02293; NID:g7922; PIDN:CAA26157.1; PID:g929565
C;Comment: This sequence is tentative because the introns have not been identified.
C;Genetics:
A;Gene: FlyBase:Egfr
A;Cross-references: FlyBase:FBgn0003731
A;Map position: 2.57F
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
F;1-732/Domain: extracellular #status predicted <EXT>
F;733-764/Domain: transmembrane #status predicted <TMM>
F;765-1330/Domain: intracellular #status predicted <INT>
F;808-1072/Domain: protein kinase homology <KIN>
F;816-824/Region: protein kinase ATP-binding motif
F;132,300,324,363,518,688,695,700/Binding site: carbohydrate (Asn) (covalent) #status pr
F;774/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F;843/Active site: Lys #status predicted
F;1181/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status predict

Query Match 24.1%; Score 1650.5; DB 1; Length 1330;
Best Local Similarity 30.2%; Pred. No. 1.8e-60;
Matches 412; Conservative 175; Mismatches 406; Indels 371; Gaps 39;

Qy	101	LEQLRIVRGTHLP-----ENYALAVLDNGDPLNNTPVTGASPGGLRELQRLSLEIL 155
		: : : : :
Db	53	LSYLRLQIRGRTLFLSVEEEKYALFV-----TSKMVLTLEIPDLRDVL 96
		: : : : :
Qy	156	KGGVLIQRNPOLCYQDTILWKDIIPHKNQLATLIDTNRSRACHPCSPCKSGSRCWGESS 215
		: : : : :
Db	97	NQGVGFHNNYNLCHWRITLQNSEIIVSNGTDAYNYDFTAPERECPKCHESCTHG-CWGEGP 155
		: : : : :
Qy	216	EDCQSRLTRTVCAAGCA--RCRKGPLPTDCHEOCAAGCTGPKHSDDCIACLHFHNHSGICELH 273
		: : : : :
Db	156	KNCQKFSLKTCSPQCAGGCRCYGPKPRECHLFCAGGTGPTOKDKCIACKNFDFEAVSKEE 215
		: : : : :
Qy	274	CPALVTYNTDTFESMPNPEGRTTFGASCVTACPNYVLTSDVSCTLVCPLNHOEVTAEDG 333
		: : : : :
Db	216	CPPMRKYNPTTYVLETNPKEGYAYGATCYKECP-GHLLRDNACVCVRSCPDKNMDKGGE-- 272
		: : : : :
Qy	334	TQRECKSKPCARVCYGLGMHLREVRAVTSANIQEFACKKIIFGSALFAPESFDG--DP 391
		: : : : :
Db	273	---CVPCNGCPKTCPCGVTVLH-----AGNISFRNCTVIDGNIRILDQTFSFGQDV 321
		: : : : :
Qy	392	ASNLA-----PLOEQVQVFETLBEITGYLYISAWPDSLPLDSVPQNLOVIGRIHLNHG 445
		: : : : :
Db	322	YANYTMGPRIYPLDPERREVSTVKYEITGYLNIETHFPQFNLSYFRNLETIHGRQLMES 381
		: : : : :
Qy	446	AY-SLTLOGLGISMGLRSLRGLGSLALIHHNTHLCFFVHTVPWDOLFNRPHQALLHTAN 504
		: : : : :
Db	382	MFAALAIVKSSLSYLEMENLKQISSGSVVIOHNRDLCCVYSNIRWPPIAQKEPEQKVWNEN 441
		: : : : :
Qy	505	RPEDEC----- 510
		: : : : :
Db	442	LRLDCLGKFLTILSVQHNIIMHI FAICREKNWHLLGSVQRGLLGSWHGSPVYLQBLQF 501
		: : : : :
Qy	511	----- 510
		: : : : :

Db 502 QMHLRLWLXIQVINSITQDSNEHQLTDACYSPSVFTSUTIERARYAIQASGLAMELE 561
Qy 511 -----VGEGLA-----CHQLCARGHC 526
Db 562 QITARSASMRHSKTLPAEGROVPRWFLGVGCASARAGIAEPLAGRAVCRKCHPCELCCTN 621
Qy 527 WGPPTQCVNCQSFRLGQECVEBCRVLQGLPREYV---NARHCLPCHPECPQNGSVTCF 583
Db 622 YGYHEQVCKSKTHYKRRQCECTEC-----PADHYTDEQRECFORHPEC-----NG---CT 669
Qy 584 GPEADQCACAHYK-----DPPF-----CVARCPGVK-PDLSYMPIWKFPDEEGAC 629
Db 670 GPGADDCSKCRNFKLFDANETGPYVNSTFMFNCTSKCPLMRHVNYQYTAIGPY-----C 723
Qy 630 QPCPINCTHSCVDLDDKCPAEQASPLTSIVSAVVGILLVVLGVFGI-LIKRRQOKI 688
Db 724 AASPFRSKITANLD-----VNMFIITGAVLVPTICILCVYIICRQOKA 770
Qy 689 RKYT--MRRLLQETELVEPLTPSGAMPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWI 746
Db 771 KKETVKTMTALSGREDSEPLRPSNIGANLCKLRIVKDAELRKGVLGMAFGFVYKGVWV 830
Qy 747 PDGENVKIPVAIKVRENTSPKANKEILDEAYVMAGVSPVSRLLGICLTSTVQLVTOL 806
Db 831 PEGENVKIPVAIKELLKSTGAESSEEFRLREAYIMASEEHVNLKLLAVCMSSQWMLITOL 890
Qy 807 MPYGCLLDHVRENRLGSLQDNLNWCQIAKMSYLEVDVLRVHRDLAARNVLVK---SPN 863
Db 891 MPLCCLLDYVNNRDKTGSKALLNWSQIAKMSYLEEKRLVHRDLAARNVLVRLLAGED 950
Qy 864 HVKITDGLARLLDIDETEHADGGKVPKIMWALESILRRRFTHQSDVMSYGVTVWELMT 923
Db 951 H---DFGLAKLLSSDSNEYKAAGKMPKWLALCIRNRVFTSKSDVWAFGVTIWELLT 1006
Qy 924 FGAKPYDGIPARETPDLLEKERLPPOPPICTIDVYIMVCKWMDSECRPRFRELVSFS 983
Db 1007 FGQRPHENIPAKDIPDLIEVGLKLEQPEICSLDYCTLLSCWHLDAAMRPTFKQLTVFA 1066
Qy 984 RMARDPORFVVIQNEGLG--PASPLDSTFYRSLLEDD---DMGLVDVAEYLVFQCGFFC 1038
Db 1067 EFADPGRYLAIGDKKTRLPA-----YTSQEKDLIRKLAPTTDGEAIAKPPDYLO 1119
Qy 1039 PDPAAGAGVHHRRHSSTRSKSGGDLITLGLPSEEEAP-----RGLAPSEAGAGSDVF 1092
Db 1120 PKAALGPS-----HRTDCT-----DEMPKLNRYCKDPNKNKNSSTGDDER 1158
Qy 1093 DG---DLGMAAKGLQSLPHDPSLQRYSEDPTVPLPSETDGVYAPLTCSPQPEYVNOP 1149
Db 1159 DSSAREVGVGNLR-----LDLPVDEDDYIMP--TCQPGENNNNM 1196
Qy 1150 DVRPQPPSPREGPLPAARPAATLERAKTLSPGKNGVVKVDFAFGGAVENPEYL---TP 1205
Db 1197 N-----NPNQNNMAAGVAAGYM-----DLIGVPVSDNPEYLLNAQTL 1235
Qy 1206 QGGAAPQPH-----PPAFSP-AFDNLYYWD 1230
Db 1236 GVGESPIPTQIGIPVGGPGTMEVKVMPGSEPTSSDHEYND 1279

RESULT 14

S35745

protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus

C:Species: avian erythroblastosis virus

C:Date: 03-Mar-1994 #sequence_revision 26-May-1995 #text_change 28-Feb-1997

C:Accession: S35745

R;Vennstrom, B.

submitted to the EMBL Data Library, March 1993

A:Reference number: S35743

A:Accession: S35745

A:Molecule type: DNA

A:Residues: 1-544 <VEN>

A:Cross-references: EMBL:X12707

C:Genetics:

A:Gene: erbB

C:Superfamily: epidermal growth factor receptor; protein kinase homology

C:Keywords: Atp; oncogene; phosphotransferase; transforming protein; tyrosine-specific p

F;135-400/Domain: protein kinase homology <KIN>

F;143-151/Region: protein kinase ATP-binding motif

F;170/Active site: Lys #status predicted

Query Match 24.1%; Score 1647; DB 2; Length 544;
Best Local Similarity 54.9%; Pred. No. 1.1e-60;
Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15;
Qy 584 GPEADQCACAHYKDPFCFVARCPGSKVPDLSYMPIWKFPDEEGACQPCPINCTHSCVDL 643
Db 1 GP--DHCMKCAHFTDGPCHVKACPAAGVLGENDTL-VMKYADANAVACQLCHPNCNTRGCKGP 57
Qy 644 DDKGCPEAQASPLTSIVSAVVGILLVVLGVFGI-LIKRRQOKIKRYTMRRLLQETEL 702
Db 58 GLEGCP---NGSKTPSIAAGVVGGLCLVVGIGLYLRRR-HIVRRKTLRRLQREEL 113
Qy 703 VEPLTPSGAMPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLR 762
Db 114 VEPLTPSGEAPNQAHLRLKETEFKVKVLGFGAFGTVYKGLWIPGEKVTIPVAIKELR 173
Qy 763 ENTSPKANKEILDEAYVMAGVSPVSRLLGICLTSTVQLVTOLMPYGCLLDHVRENRR 822
Db 174 EATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYIREHKDN 233
Qy 823 LGSQDILLNWCQIAKMSYLEVDVLRVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETE 882
Db 234 IGSQYLLNWCQIAKMSYLEERHMHVHRDLAARNVLVKTPHQVKITDFGLAKQLGADEKE 293
Qy 883 YHADGGKVPKIMWALESILRRRFTHQSDVMSYGVTVWELMTFFGAKPYDGI-PAREIPDLLE 942
Db 294 YHAEKGKVPKIMWALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGI-PASISSVLE 353
Qy 943 KGERLPPOPPICTIDVYIMVCKWMDSECRPRFRELVSFMRMARDPORFVVIQ-NBDLG 1001
Db 354 KGERLPPOPPICTIDVYIMVCKWMSDADSRPKRELIAEFKWARDPPRYLVIOGDERMH 413
Qy 1002 PASPLDSTFYRSLLEDDMGDLVDAAEYLVFQCGFFCDDPAPGAGGVHHRHSSTRSRG 1061
Db 414 LPSPTDSKFRYTLMEEDMEDIVDAEYLVPHQGF-----NSPST--- 454
Qy 1062 GGDLITGLPSEEEAPRSL-----APSEGAGSDVFDGDLGMAAKGLQSLPHDPSLQ 1116
Db 455 -----SRTPLLSLSATSNNATNCIDRNG-----H----- 481
Qy 1117 RYSEDPTVPLPSETDGVYAPLTCSPQPEYVNOPDVRPQPPSPREGPLPAARPAAT-LER 1175
Db 482 -----FVREDGFL-----PAPEYVNO--LMPKPSAMVQNIYVILSLTAISK 523
Qy 1176 AKTILSPKNGVVKDVFAGGAVENPEYL 1203
Db 524 LPIDSRYN-----SHSTAVDNPEYL 544

RESULT 15

S00727

kinase-related transforming protein (erbB) (EC 2.7.1.-) - avian erythroblastosis virus (

C:Species: avian erythroblastosis virus

C:Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 28-Feb-1997

C:Accession: S00727

R;Scotting, P.; Vennstrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.

Oncogene Res. 1, 265-278, 1987

A:Title: Common site of mutation in the erbB gene of avian erythroblastosis virus mutant

A:Reference number: S00727; MUID:88217326; PMID:2897102

A:Accession: S00727

A:Molecule type: DNA

A:Residues: 1-545 <SCO>

A:Cross-references: EMBL:X06943

C:Genetics:

A:Gene: erbB

Search completed: July 22, 2003, 09:08:10
Job time : 31.0157 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:12:49 ; Search time 10.2793 Seconds
(without alignments)
5088.033 Million cell updates/sec

Title: SEQ4-59-73-14

Perfect score: 6847

Sequence: 1 MELAALCRWGLLALLPPGA.....TFKGTPTAENPEYLGLDVPV 1261

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6707	98.0	1255	1	ERB2_HUMAN
2	5907	86.3	1257	1	ERB2_RAT
3	5901.5	86.2	1254	1	ERB2_MESAU
4	3096	45.2	1210	1	EGFR_HUMAN
5	3075	44.9	1210	1	EGFR_MOUSE
6	2930.5	42.8	1308	1	ERB4_HUMAN
7	2911	42.5	1308	1	ERB4_RAT
8	2649.5	38.7	1167	1	XMRK_XIPPA
9	2382.5	34.8	1342	1	ERB3_HUMAN
10	2316.5	33.8	1339	1	ERB3_RAT
11	1919	28.0	1426	1	EGFR_DROME
12	1749.5	25.6	634	1	ERBB_ALV
13	1703	24.9	604	1	ERBB_AVIER
14	1630	23.8	540	1	ERBB_AVIEU
15	1555	22.7	703	1	EGFR_CHICK
16	1274	18.6	1323	1	LT23_CABEL
17	1142.5	16.7	245	1	ERB2_MOUSE
18	736	10.7	1363	1	ILPR_BRALA
19	718	10.5	1382	1	INSR_HUMAN
20	713	10.4	1383	1	INSR_RAT
21	712.5	10.4	1372	1	INSR_MOUSE
22	706	10.3	1300	1	IRR_MOUSE
23	698	10.2	1297	1	IRR_HUMAN
24	692.5	10.1	1300	1	IRR_CAVPO
25	691	10.1	1607	1	MLPR_LYNMT
26	670	9.8	1477	1	HTK7_HYDAT
27	669	9.8	1367	1	IG1R_HUMAN
28	659	9.6	1373	1	IG1R_MOUSE
29	655.5	9.6	1370	1	IG1R_RAT
30	623	9.1	1390	1	INSR_AEDAE
31	613	9.0	987	1	ERB4_HUMAN
32	613	9.0	2146	1	INSR_DROME
33	595	8.7	987	1	EPB4_MOUSE

RESULT 1

ID	ERB2_HUMAN	STANDARD;	PRT;	1255 AA.
AC	P04626;			
DT	13-AUG-1987 (Rel. 05, Created)			
DT	13-AUG-1987 (Rel. 05, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)			
DE	(p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell surface receptor HER2) (MLN 19).			
DE	ERBB2 OR HER2 OR NGL OR NEU.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=86118663; PubMed=3003577;			
RA	Yanamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,			
RA	Saito T., Toyoshima K.;			
RT	"Similarity of protein encoded by the human c-erb-B-2 gene to			
RT	epidermal growth factor receptor.";			
RL	Nature 319:230-234(1986).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=86070181; PubMed=2999974;			
RA	Coussens L., Yang-Peng T.L., Liao Y.C., Chen E., Gray A.,			
RA	McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J.,			
RA	Francke U., Levinson A., Ullrich A.;			
RT	"Tyrosine kinase receptor with extensive homology to EGF receptor			
RT	shares chromosomal location with neu oncogene.";			
RL	Science 230:1132-1139(1985).			
RN	[3]			
RP	SEQUENCE OF 737-1031 FROM N.A.			
RX	MEDLINE=86016729; PubMed=2995967;			
RA	Semba K., Kamata N., Toyoshima K., Yamamoto T.;			
RT	"A v-erbB-related protooncogene, c-erbB-2, is distinct from the			
RT	c-erbB-1/epidermal growth factor-receptor gene and is amplified in a			
RT	human salivary gland adenocarcinoma.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).			
RN	[4]			
RP	VARIANTS VAL-654 AND VAL-655.			
RX	MEDLINE=93194196; PubMed=8095488;			
RA	Ehsani A., Low J., Wallace R.B., Wu A.M.;			
RT	"Characterization of a new allele of the human ERBB2 gene by allele-			
RT	specific competition hybridization.";			
RL	Genomics 15:426-429(1993).			
CC	-1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,			
CC	ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A			
CC	POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-			
CC	ALPHA AND AMPHIREGULIN.			
CC	-1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein			
CC	tyrosine phosphate.			
CC	-1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS			
CC	(POTENTIAL).			
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein.			

Q07494 gallus gall
Q03145 mus musculu
P07949 homo sapien
P09759 rattus norv
P29317 homo sapien
P54762 homo sapien
Q91736 xenopus lae
Q91571 xenopus lae
Q00944 gallus gall
Q91738 xenopus lae
P53356 hydra atten
P34132 mus musculu


```
QY 721 LKETLRKVKVLGSGAGFTVYKGIWI PDGENVKIPVAIKVLRNTPSKANKEIIDEAYVM 780
DQ 715 LKETLRKVKVLGSGAGFTVYKGIWI PDGENVKIPVAIKVLRNTPSKANKEIIDEAYVM 774
QY 781 AGVGSPPVSRLLGICLTSTVOLVTQMLPYGCLLDHVRNRCRLGSDQLLNWCMIAGKMS 840
DQ 775 AGVGSPPVSRLLGICLTSTVOLVTQMLPYGCLLDHVRNRCRLGSDQLLNWCMIAGKMS 834
QY 841 YLEDRVLRDLAARNVLKSPNHVKITDFGLARLLDDIDEYHADGKVPKIMWALESI 900
DQ 835 YLEDRVLRDLAARNVLKSPNHVKITDFGLARLLDDIDEYHADGKVPKIMWALESI 894
QY 901 LRRRFTHQSDVMSYGVTVWELMTGAKPYDGIPIAREIPDLEKGERLPQPPICITDVYMI 960
DQ 895 LRRRFTHQSDVMSYGVTVWELMTGAKPYDGIPIAREIPDLEKGERLPQPPICITDVYMI 954
QY 961 MVKCMWIDSECRPRELVSFEFSWARDPQRFVVIQNEIDLGPAFLDSTFYRSLLDDDDM 1020
DQ 955 MVKCMWIDSECRPRELVSFEFSWARDPQRFVVIQNEIDLGPAFLDSTFYRSLLDDDDM 1014
QY 1021 GDLVDAEYLVPQOGFCPPDAPGAGGVMVHRHRSSTRSGGDLTLGLEPSEEEAPRSP 1080
DQ 1015 GDLVDAEYLVPQOGFCPPDAPGAGGVMVHRHRSSTRSGGDLTLGLEPSEEEAPRSP 1074
QY 1081 LAPSEGAGSDVFDGDLGWAAGKGLQSLPTHDPSPLOQYSEDPTVPLPSETDGYVAPLTCS 1140
DQ 1075 LAPSEGAGSDVFDGDLGWAAGKGLQSLPTHDPSPLOQYSEDPTVPLPSETDGYVAPLTCS 1134
QY 1141 POPEYVNPQDVRPOPPSPREGPLPAARPAAGATLERAKTLPSPKNGVVKDVPFAFGGAVENP 1200
DQ 1135 POPEYVNPQDVRPOPPSPREGPLPAARPAAGATLERAKTLPSPKNGVVKDVPFAFGGAVENP 1194
QY 1201 EYLTPQGGAAPOPHPPPAFSPAFNLVYWDQDPPERGAPPTSTKGTPTAENPEYVLGLDVP 1260
DQ 1195 EYLTPQGGAAPOPHPPPAFSPAFNLVYWDQDPPERGAPPTSTKGTPTAENPEYVLGLDVP 1254
QY 1261 V 1261
DQ 1255 V 1255

RESULT 2
ERB2 RAT STANDARD; PRT; 1257 AA.
AC P06494;
DT 01-JAN-1988 (Rel. 06, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2). (NEU proto-oncogene) (C-erbB-2) (Epidermal growth factor
DE receptor-related protein).
DE ERBB2 OR NEU.
GN Rattus norvegicus (Rat).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Neuroblastoma;
RX MEDLINE=86118662; PubMed=3945311;
RA Bargmann C.I., Hung M.-C., Weinberg R.A.;
RT "The neu oncogene encodes an epidermal growth factor receptor-related
RT protein.";
RL Nature 319:226-230(1986).
RN [2]
RP SEQUENCE OF 852-905 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=91222560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system.";
RL Neuron 6:691-704(1991).
```

```

RN [3]
RP STRUCTURE BY NMR OF 650-668.
RX MEDLINE=92155181; PubMed=1346763;
RA Gullick W.J., Bottomley A.C., Loftis F.J., Doak D.G., Mulvey D.,
RA Newman R., Crumpton M.J., Sternberg M.J.E., Campbell I.D.;
RT "Three dimensional structure of the transmembrane region of the proto-
RT oncogenic and oncogenic forms of the neu protein.";
RL EMBO J. 11:43-48(1992).
CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS.
CC THE CONSTITUTIVELY ACTIVATED ONCOGENIC VARIANT FORMS A HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- PM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X03362; CAA27059.1; ALT_INIT.
CC PIR; A24562; TVRTNU.
CC HSP; P11362; IFGK.
CC InterPro; IPR000494; EGFR_L_domain.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR001245; Tyr_pkinase.
CC InterPro; IPR004019; YLP_motif.
CC Pfam; PF00069; pkinase; 1.
CC Pfam; PF00757; Furin-like; 1.
CC Pfam; PF01030; Recep_L_domain; 2.
CC Pfam; PF02757; YLP; 2.
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00261; FU; 3.
CC SMART; SM00219; TyrKc; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
CC Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
CC Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
CC Proto-oncogene; Disease mutation.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 1257 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
FT DOMAIN 22 654 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 655 677 POTENTIAL.
FT DOMAIN 678 1257 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 159 369 CYS-RICH.
FT DOMAIN 473 646 CYS-RICH.
FT DOMAIN 722 989 PROTEIN KINASE.
FT NP_BIND 728 736 ATP (BY SIMILARITY).
FT BINDING 755 755 ATP (BY SIMILARITY).
FT ACT_SITE 847 847 BY SIMILARITY.
FT DISULFID 196 205 BY SIMILARITY.
FT DISULFID 200 213 BY SIMILARITY.
FT DISULFID 221 228 BY SIMILARITY.
FT DISULFID 225 236 BY SIMILARITY.
FT DISULFID 237 245 BY SIMILARITY.
FT DISULFID 241 253 BY SIMILARITY.
FT DISULFID 256 265 BY SIMILARITY.
FT DISULFID 269 295 BY SIMILARITY.
FT DISULFID 300 312 BY SIMILARITY.
FT DISULFID 316 332 BY SIMILARITY.
FT DISULFID 335 339 BY SIMILARITY.
FT DISULFID 513 522 BY SIMILARITY.
```

FT DISULFID 517 530 BY SIMILARITY.
FT DISULFID 533 542 BY SIMILARITY.
FT DISULFID 546 562 BY SIMILARITY.
FT DISULFID 565 578 BY SIMILARITY.
FT DISULFID 569 586 BY SIMILARITY.
FT DISULFID 589 598 BY SIMILARITY.
FT DISULFID 602 625 BY SIMILARITY.
FT DISULFID 628 636 BY SIMILARITY.
FT DISULFID 632 644 BY SIMILARITY.
FT MOD RES 1141 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD RES 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 188 188 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 260 260 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 631 631 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).
SQ SEQUENCE 1257 AA; 138831 MW; 6129264583011402 CRC64;

Query Match 86.3%; Score 5907; DB 1; Length 1257;
Best Local Similarity 86.3%; Pred. No. 2.2e-307;
Matches 1096; Conservative 48; Mismatches 104; Indels 22; Gaps 5;

Qy 1 MELAALCRWGLLLALLPPGAASCTCTGDMKRLPASPETHDMLRHLRYOGCQVVOGNL 60
Db 1 MELAAWCRWGLLLALLPPGIAGTQVCTGDMKRLPASPETHDMLRHLRYOGCQVVOGNL 60

Qy 61 ELTYLPTNASLSFLQDIQEVQFNFTVSFWL-----RVKVSASHLEQLRIVRGTL 113
Db 61 ELTYVPAASLSFLQDIQEVQ-----GYMLIAHNVKRVPL-----QLRIVRGTL 107

Qy 114 FEDNYALAVLDNGDPLNNTTPTV-GASPGGLRELQLSLRLEILKGGVLIQRNPOLCYQDT 172
Db 108 FEDYALAVLDNRDPQDNVAASSTGRTPEGLRELQLSLRLEILKGGVLIQRNPOLCYDM 167

Qy 173 ILWKDIHFNKNQALTLTIDNRACHPCSPCMKGRCSWSSSDCSLRTVCAGGCR 232
Db 168 VLWKDVFRKNQALPVDIDNRACHPCPACPKDNHCWGESPEDCILLTCTISGCR 227

Qy 233 CKGLPLTDCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNE 292
Db 228 CKGLPLTDCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNE 287

Qy 293 GRYTFGASCTVACPNYLSLTDVSGCTVCPHNOEVTAEADGTQCEKSCPCARVCYGLG 352
Db 288 GRYTFGASCTVTCPNYLSLTDVSGCTVCPHNOEVTAEADGTQCEKSCPCARVCYGLG 347

Qy 353 MEHLREVRVTSANTIQEPAGCKKIFGSLAFIPESFDGDPASNTAPLOEQLOVFTLEEI 412
Db 348 MEHLRGARVTSNDVQEPDGCKKIFGSLAFIPESFDGDPSSGIAPLPEQLOVFTLEEI 407

Qy 413 TGYLYISAWPDSLPDLSVFQNLQVIRGILHNGAYSLLTQGLGSIWGLRSLRGLSGLA 472
Db 408 TGYLYISAWPDSLRDLVSFQNLRIIRGILHNGAYSLLTQGLGSIHSLRSLRGLSGLA 467

Qy 473 LIHNTHTLCFVHTVPMDLFRNPQALLHTANREDE-CVGEGLACHQLCARGHCWGP 531
Db 468 LIHRTNAHLCFVHTVPMDLFRNPQALLHNGRNEEDLCVSSGLVNCSLCAHGCWGP 527

Qy 532 TQCVNCSOFRLGQECVECRVQLGLPREYNVNRHCLPCHPCQPNQSGVTCFGEADQCV 591
Db 528 TQCVNCSHFLRGQECVECRVWGLPREYVSDKCLPCHPCQPNQSGVTCFGEADQCA 587

Qy 592 ACAHYKDPFPFCVAPCPGKPDLSYMPITWKPDPDEGACQPCPINCTHSCVDLDKGCPAE 651
Db 588 ACAHYKDSKSCVAPCPGKPDLSYMPITWKPDPDEGACQPCPINCTHSCVDLDKGCPAE 647

Qy 652 QRASPLTSIVAVVGLVVLVGVGLIKRQOKIRKYMRRLQETLVEPLTPSGA 711
Db 648 QRASPVTFIATVGVVGLVVLVGVGLIKRQOKIRKYMRRLQETLVEPLTPSGA 707

Qy 712 'MPNOAQMRLKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANK 771

Db 708 'MPNOAQMRLKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANK 767
Qy 772 EILDEAYVWAGVSPVYVRLGLGICLTSTVOLVTQLMPYGCGLLDHVRNRLGASQDILNW 831
Db 768 EILDEAYVWAGVSPVYVRLGLGICLTSTVOLVTQLMPYGCGLLDHVRNRLGASQDILNW 827
Qy 832 CMQIAGMSVLEDRVLRVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVP 891
Db 828 CVQIAGMSVLEDRVLRVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVP 887
Qy 892 IKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPP 951
Db 888 IKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPP 947
Qy 952 ICTIDVIMVCKWMDIDSECRPRRELVSFSESMARDPQRFVVIQNEIDLGPASPLDSTFY 1011
Db 948 ICTIDVIMVCKWMDIDSECRPRRELVSFSESMARDPQRFVVIQNEIDLGPSSPMDSTFY 1007
Qy 1012 RSLLEDDMGDLVDAEYLVPOQGFPCDPAPCAGGVMVHRRSSSTRSGGDLTLGLEP 1071
Db 1008 RSLLEDDMGDLVDAEYLVPOQGFSPDPTGTGTAHRRHSSSTRSGGDLTLGLEP 1067
Qy 1072 SESEAPRSLAPSEAGSDVDFDGLMGAAKGLQSLPTHDPSPLOQRYSEDTVPPLPSETD 1131
Db 1068 SESEAPRSLAPSEAGSDVDFDGLMGVTKGLQSLSPHDLSPLOQRYSEDTVPPLPETD 1127
Qy 1132 GYVAPITCSQPSQPVYVQNPVVRPQPPPREGPLPAAPAGATLERAKTSLSPKNGVWQVDF 1191
Db 1128 GYVAPLACSPQPSYVQNPVVRPQPPPREGPLPAAPAGATLERAKTSLSPKNGVWQVDF 1187
Qy 1192 AFGSAVENPEYLTPOQGAAPQPPHPPAFSPAFDNLVYWDODPPERGAPPTFEGTPTAEN 1251
Db 1188 AFGSAVENPEYLTPOQGAAPQPPHPPAFSPAFDNLVYWDODPPERGAPPTFEGTPTAEN 1247
Qy 1252 PEYLGLDVVP 1261
Db 1248 PEYLGLDVVP 1257

RESULT 3
ERB2 MESAU
ID ERB2 MESAU STANDARD; PRT; 1254 AA.
AC Q6053;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2).
GN ERB2 OR NEU
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Nerve;
RX MEDLINE=94193007; Pubmed=7908275;
RA Nakamura T., Ushijima T., Ishizaka Y., Nagao M., Arai M.,
RA Yamazaki Y., Iehikawa T.;
RT "Cloning and activation of the Syrian hamster neu proto-oncogene.";
RL Gene 140:251-255(1994).
CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE

Qy 1020 MGDLDVDAEYLVPOQGFPPDPAGAGMGVHRRSSSTRSGGDLTLGLEPSEEEAPRS 1079
 Db 1014 MGDLDVDAEYLVPOQGFPPDPAGAGMGVHRRSSSTRSGGDLTLGLEPSEEEAPRS 1073
 Qy 1080 PLAPSEGAGSDVFDGLGMGAAGLQSLPTHDPSPLOQYSEDPTVPLPSETDGYVAPLTC 1139
 Db 1074 PLAPSEGAGSDVFDGELGMGATKGPQSI SPRDLSPLQYSEDPTLPLPTETDGYVAPLAC 1133
 Qy 1140 SPQPEYNQPVPRQPPSPREGPIPLAARPAAGATLERAKTLSPGKGVGVKQVFAFGGAVEN 1199
 Db 1134 SPQPEYNQPVPRQPPSPREGPIPLAARPAAGATLERAKTLSPGKGVGVKQVFTFGGAVEN 1193
 Qy 1200 PEYLTPOGGAAPQHPHPPAFSPADNLYWQDPPERGAPSTFKGPTAENPEYVLGLDV 1259
 Db 1194 PEYLVPRGSGSQPH-PPALCPAFDNLVWQDPSERGSPPNTFEGPTAENPEYVLGLDV 1252
 Qy 1260 PV 1261
 Db 1253 PV 1254

RESULT 4
 EGFR HUMAN STANDARD; PRT; 1210 AA.
 AC P00533; P06268; Q14225; Q9UMD7; Q9UMD8; Q9UMG5; Q92795; Q00732;
 AC O00688; Q9B2S2; Q9H2C9; Q9GZX1; Q9H3C9;
 DT 21-JUN-1986 (Rel. 01, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor
 protein-tyrosine kinase ErbB-1).
 GN EGFR OR ERBB1
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=84219729; PubMed=63289112;
 RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
 RA Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J.,
 RA Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;
 RT "Human epidermal growth factor receptor cDNA sequence and aberrant
 RT expression of the amplified gene in A431 epidermoid carcinoma cells.";
 RL Nature 309:418-425(1984).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX TISSUE=Placenta;
 RC MEDLINE=95382957; PubMed=7654368;
 RA Ilekis J.V., Stark B.C., Scoccia B.;
 RT "Possible role of variant RNA transcripts in the regulation of
 RT epidermal growth factor receptor expression in human placenta.";
 RL Mol. Reprod. Dev. 41:149-156(1995).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX TISSUE=Placenta;
 RC MEDLINE=97078686; PubMed=8918811;
 RA Reiter J.L., Mailhe N.J.;
 RT "A 1.8 kb alternative transcript from the human epidermal growth
 RT factor receptor gene encodes a truncated form of the receptor.";
 RL Nucleic Acids Res. 24:4050-4056(1996).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX TISSUE=Placenta;
 RC MEDLINE=97265457; PubMed=9103388;
 RA Ilekis J.V., Gariti J., Niederberger C., Scoccia B.;
 RT "Expression of a truncated epidermal growth factor receptor-like
 RT protein (TEGFR) in ovarian cancer.";
 RL Gynecol. Oncol. 65:36-41(1997).
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).
 RC TISSUE=Placenta;
 RX MEDLINE=21100872; PubMed=11161793;

RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
 RA Schehl Sinclair C., Pearshall R.S., Green P.J., Yee D., Lampland A.L.,
 RA Balasubramaniam S., Crossley T.D., Magnuson T.R., James C.D.,
 RA Mailhe N.J.;
 RT "Comparative genomic sequence analysis and isolation of human and
 RT mouse alternative EGFR transcripts encoding truncated receptor
 RT isoforms.";
 RL Genomics 71:1-20(2001).
 RN [6]
 RP SEQUENCE OF 575-687 FROM N.A.
 RA Reiter J.L., Threadgill D.W., Danielsen A.J., Schehl C.M., T.R.,
 RA Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
 RA Mailhe N.J.;
 RT "Human and mouse alternative EGFR transcripts encoding only the
 RT extracellular domain of the receptor.";
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP SEQUENCE OF 713-924 FROM N.A.
 RX MEDLINE=84196372; PubMed=6326261;
 RA Lin C.R., Chen W.S., Krueger W., Stolarsky L.S., Weber W.,
 RA Evans R.M., Verma I.M., Gill C.N., Rosenfeld M.G.;
 RT "Expression cloning of human EGF receptor complementary DNA: gene
 RT amplification and three related messenger RNA products in A431
 RT cells.";
 RL Science 224:843-848(1984).
 RN [8]
 RP SEQUENCE OF 150-962 FROM N.A.
 RX MEDLINE=84245835; PubMed=6330563;
 RA Xu Y.H., Ishii S., Clark A.J., Sullivan M., Wilson R.K., Ma D.P.,
 RA Roe B.A., Merlino G.T., Pastan I.;
 RT "Human epidermal growth factor receptor cDNA is homologous to a
 RT variety of RNAs overproduced in A431 carcinoma cells.";
 RL Nature 309:806-810(1984).
 RN [9]
 RP SEQUENCE OF 1028-1210 FROM N.A.
 RX MEDLINE=85046483; PubMed=6093780;
 RA Simmen F.A., Gope M.L., Schulz T.Z., Wright D.A., Carpenter G.,
 RA O'Malley B.W.;
 RT "Isolation of an evolutionarily conserved epidermal growth factor
 RT receptor cDNA from human A431 carcinoma cells.";
 RL Biochem. Biophys. Res. Commun. 124:125-132(1984).
 RN [10]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=88217333; PubMed=3329716;
 RA Haley J.D., Whittle N., Bennett P., Kinchington D., Ullrich A.,
 RA Waterfield M.D.;
 RT "The human EGF receptor gene: structure of the 110 kb locus and
 RT identification of sequences regulating its transcription.";
 RL Oncogene Res. 1:375-396(1987).
 RN [11]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=91107677; PubMed=1988448;
 RA Haley J.D., Waterfield M.D.;
 RT "Contributory effects of de novo transcription and premature
 RT transcript termination in the regulation of human epidermal growth
 RT factor receptor proto-oncogene RNA synthesis.";
 RL J. Biol. Chem. 266:1746-1753(1991).
 RN [12]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=85270438; PubMed=2991899;
 RA Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;
 RT "Characterization and sequence of the promoter region of the human
 RT epidermal growth factor receptor gene.";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).
 RN [13]
 RP SEQUENCE OF 540.
 RA Kohda D.;
 RT Submitted (SEP-1997) to the SWISS-PROT data bank.
 RN [14]
 RP RECEPTOR ACTIVITY.
 RX MEDLINE=84191554; PubMed=6325948;
 RA Mroczkowski B., Mosig G., Cohen S.;
 RT "ATP-stimulated interaction between epidermal growth factor receptor

QY 964 CWMIDSECRPRFELVSEFMRWDPQRFVVIQ-NEDLGPASPLDSTFYRLSDDDDMGD 1022
 DB 950 CWMIDADSRPRFELVSEFMRWDPQRFVVIQ-ODERWHLSPDTSFYRLMDEEDMD 1009
 QY 1023 LVDAEYLVPQGGFFCPDPAFCAGGMVHRRHSRSTGSGGDLTLGLPESEEAAPRSPLA 1082
 DB 1010 VVDAEYLVPQGGFF-----SSPSTSRTPLL 1035
 QY 1083 PSEGAGSDVFGDGLMGAAKGLSLPHDPSPLQRYSEDPVPLPSET--DGYVAPLTC 1140
 DB 1036 SSLSATSN--NSTVACIDRNLQCPKEKEDSFLQRYSSDPTGALTEDSIDTFL----- 1087
 QY 1141 POPEVNVQDVRPOPSPREGPLPAARPAAGATLERAKTLSPKNGVVKDVFAGGAVENP 1200
 DB 1088 PVPEYINO-SYKPRAGSVQNVHNLPLN-----AFSRPHYQD--PHSTAVGNP 1136
 QY 1201 EYL-TPQGAAPQHPPPPAFAFDNLVYWDQ-----DP-----PERGAPSTF 1243
 DB 1137 EYLVNTVQ-----PTCVNSTFDSPAHWAQKSHQISLNDPQYQDDFFPKPEAKPNGIF 1187
 QY 1244 KGTPTAENPEYL 1255
 DB 1188 KGS-TAENAEYL 1198
 RESULT 5
 EGFR_MOUSE
 ID EGFR_MOUSE STANDARD; PRT; 1210 AA.
 AC Q01279;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Epidermal growth factor receptor precursor (EC 2.7.1.112).
 GN EGFR.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c; TISSUE=Liver;
 RX MEDLINE=93026370; PubMed=1408137;
 RA Avivi A., Skorecki K., Yayon A., Givol D.;
 RT "Promoter region of the murine fibroblast growth factor receptor 2
 (bek/KGFR) gene.";
 RL Oncogene 7:1957-1962(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c, and CD-1; TISSUE=Liver, and Decidia;
 RX MEDLINE=93126380; PubMed=7678348;
 RA Paria B.C., Das S.K., Andrews G.K., Dey S.K.;
 RT "Expression of the epidermal growth factor receptor gene is regulated
 in mouse blastocysts during delayed implantation.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:55-59(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c; TISSUE=Liver;
 RA Hibbs M.L.;
 RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=B6/C3; TISSUE=Liver;
 RX MEDLINE=94170986; PubMed=8125255;
 RA Luetteke N.C., Phillips H.K., Qiu T.H., Copeland N.G., Earp H.S.,
 RA Jenkins N.A., Lee D.C.;
 RT "The mouse waved-2 phenotype results from a point mutation in the EGF
 receptor tyrosine kinase.";
 RL Genes Dev. 8:399-413(1994).
 RN [5]
 RP SEQUENCE OF 1-714 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=91232866; PubMed=2030916;
 RA Avivi A., Lax I., Ullrich A., Schlessinger J., Givol D., Morse B.;

"Comparison of EGF receptor sequences as a guide to study the ligand
 binding site.";
 RL Oncogene 6:673-676(1991).
 RN [6]
 RP SEQUENCE OF 969-1117 FROM N.A.
 RC STRAIN=C3H;
 RA Eisinger D.P., Serrero G.;
 RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF.
 CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
 CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
 CC dimerization, internalization of the EGF-receptor complex,
 CC induction of the tyrosine kinase activity, stimulation of cell DNA
 CC synthesis, and cell proliferation.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X78987; CAA55587.1; -;
 CC EMBL; U03425; AAA17899.1; -;
 CC EMBL; X59698; CAA42219.1; -;
 CC EMBL; L06864; AAA53029.1; -;
 CC EMBL; Z12608; CAA78249.1; -;
 CC HSP; P11362; IPGK.
 CC MGD; MGI:95294; Egr.
 CC InterPro; IPR000494; EGFR_L_domain.
 CC InterPro; IPR000719; Euk_pkinase.
 CC InterPro; IPR002174; Furin-like.
 CC InterPro; IPR001245; Tyr_pkinase.
 CC Pfam; PF00069; pkinase; 1.
 CC Pfam; PF00757; Furin-like; 1.
 CC Pfam; PF01030; Recep_L_domain; 2.
 CC ProDom; PD000001; Euk_pkinase; 1.
 CC SMART; SM00361; FU; 3.
 CC SMART; SM00319; TyrKc; 1.
 CC PROSITE; PS00107; PROTEIN KINASE ATP; 1.
 CC PROSITE; PS00109; PROTEIN KINASE TYR; 1.
 CC PROSITE; PS00111; PROTEIN KINASE DOM; 1.
 CC Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
 CC Tyrosine-protein kinase; ATP-binding; Phosphorylation; Repeat.
 CC SIGNAL 1 24 POTENTIAL.
 CC CHAIN 1 24 EPIDERMAL GROWTH FACTOR RECEPTOR.
 CC DOMAIN 25 647 EXTRACELLULAR (POTENTIAL).
 CC TRANSMEM 648 670 POTENTIAL.
 CC DOMAIN 671 1210 CYTOPLASMIC (POTENTIAL).
 CC REPEAT 75 300 APPROXIMATE.
 CC REPEAT 390 600 APPROXIMATE.
 CC DOMAIN 1028 1071 SER-RICH.
 CC DOMAIN 714 981 PROTEIN KINASE.
 CC NP BIND 720 728 ATP (BY SIMILARITY).
 CC BINDING 747 747 ATP (BY SIMILARITY).
 CC ACT SITE 839 839 BY SIMILARITY.
 CC DISULFID 190 199 BY SIMILARITY.
 CC DISULFID 194 207 BY SIMILARITY.
 CC DISULFID 215 223 BY SIMILARITY.
 CC DISULFID 219 231 BY SIMILARITY.
 CC DISULFID 232 240 BY SIMILARITY.
 CC DISULFID 236 248 BY SIMILARITY.
 CC DISULFID 251 260 BY SIMILARITY.
 CC DISULFID 264 291 BY SIMILARITY.
 CC DISULFID 295 307 BY SIMILARITY.
 CC DISULFID 311 326 BY SIMILARITY.
 CC DISULFID 329 333 BY SIMILARITY.

FT	NP_BIND	724	732	ATP (BY SIMILARITY).
FT	BINDING	751	751	ATP (BY SIMILARITY).
FT	ACT_SITE	843	843	BY SIMILARITY.
FT	DISULFID	189	197	BY SIMILARITY.
FT	DISULFID	193	205	BY SIMILARITY.
FT	DISULFID	213	221	BY SIMILARITY.
FT	DISULFID	217	229	BY SIMILARITY.
FT	DISULFID	230	238	BY SIMILARITY.
FT	DISULFID	234	246	BY SIMILARITY.
FT	DISULFID	249	258	BY SIMILARITY.
FT	DISULFID	262	289	BY SIMILARITY.
FT	DISULFID	293	304	BY SIMILARITY.
FT	DISULFID	308	323	BY SIMILARITY.
FT	DISULFID	326	330	BY SIMILARITY.
FT	DISULFID	503	512	BY SIMILARITY.
FT	DISULFID	507	520	BY SIMILARITY.
FT	DISULFID	523	532	BY SIMILARITY.
FT	DISULFID	536	552	BY SIMILARITY.
FT	DISULFID	555	569	BY SIMILARITY.
FT	DISULFID	559	577	BY SIMILARITY.
FT	DISULFID	580	589	BY SIMILARITY.
FT	DISULFID	593	614	BY SIMILARITY.
FT	DISULFID	617	635	BY SIMILARITY.
FT	DISULFID	621	633	BY SIMILARITY.
FT	MOD_RES	1162	1162	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES	1188	1188	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES	1258	1258	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES	1284	1284	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	CARBOHYD	138	138	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	174	174	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	181	181	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	253	253	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	358	358	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	410	410	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	473	473	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	495	495	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	548	548	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	576	576	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	620	620	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	VARSPLIC	626	648	NGPTSHDCIYPTWGHSTLTPQHA -> IGSSIEDCIGLMD (IN ISOFORM JM-B).
FT	SEQUENCE	1308 AA;	146807 MW;	554AE80985D88761 CRC64;
Qy	Query Match	42.8%;	Score 2930.5;	DB 1; Length 1308;
Db	Best Local Similarity	44.5%;	Prod. No. 8.2e-149;	
Qy	Matches 606;	Conservative 183;	Mismatches 375;	Indels 197; Gaps 30;
Qy	9	WGILLALLPGAA----	STQVCTGTDMKRLRPASPETHLDMLRHLRYOGCQVVOGNLELT	64
Db	8	WVVSLVLAAGTVQPSDSQSVCAGTENKLSLDLEQOYRALRKYIENCVEVWGNLEITS	67	
Qy	65	LPTNASLSFLQDIQEV-----	QFNFTVFSWLVRPKVKSASHLEQRILRVRGTOLE	115
Db	68	IEHRNRLSFLRSVREVTGYVLVALNGPRYLPL-----	ENLRIRGTLYE	112
Qy	116	DNYALAVLNDGDPNNITPTVTGASPGGLRELQRLSLTEILKGGVLIORNPOLCVQDITLW	175	
Db	113	DYALALAIFLNRYKDNFN-----	GLQELGKLNLTIELNGVYVDQNFCLCVADIHW	163
Qy	176	KDIFHKNNQLALTLIDTNRSRACHPCSPMCKSRMGESSEDCQSLTRTVCAGGC-ARCK	234	
Db	164	QDIVRPWPNSLTLVSTNGSGGCRCHKSCVTG-RCWGPTEHNHCQTLTRTVCAEQCDGRCY	222	
Qy	235	GPILPTDCHEQCAAGCTGPKHSDCLACLHFNHSGI CELHCPALVTYNTDTPFESHPNPEGR	294	
Db	223	GPVVSDDCHRECAAGCGSPKDTDFCACMFNFDNSGACVTCQCPQTFVYVNPPTTFOLEHFNNAK	282	
Qy	295	YTFGASCVTACPNYLSTDVGSCTLVLCPLHNQEVTAEDGTORCEKSKPCARVCYGLGME	354	
Db	283	YTTGAFCKVKCPHNFV-VDSSSCVRACPPSSKMEV-EENGIKMKCKPCTDIPCACDGICTG	340	
Qy	355	HLREVRVTSANTIOEFAGCKIKFGSLAFLPESFDGDPASNTAPIQBPOLQVFTLEBITG	414	

Db 341 SLMSAQTVDSNIDKFNCTKINGNLFLVLTGIGHDPVNAIEADPEKLVNFRVTRVREITG 400
Qy 415 YLYISAMPDLSPLSVFQNLQVIRGRILHNGAYSITLQGLGISWLGRLSRLFELGSLALI 474
Db 401 FLNIQSPNNMTDFSVSNLVTIGRVLVSLGLSLLILKQGGITSLQFSLKEISAGNIYI 460
Qy 475 HHNTHLCFVHTVPNDOLFNRPHQALLHTANRPEDECVEGSLACHOLCARGHCWPGPTQC 534
Db 461 TDNSNLCCYHTINNTLFTSTINQIRIVIRDNKKAENCTAEGMVCNHLCCSSDCGCGPGDQC 520
Qy 535 VNCQFRLRGQCBECRVLQGLPREYVNAHCLPCHPECQ-ONGSVTCFQPEADQCVCAC 593
Db 521 LSCRFSGRGRICIESCNLYDGEFREFNGSICVCEDCQCEKMEGLLTCHGPGPDNTKC 580
Qy 594 AHYKDPFPCVACPSGVKPDLSYMPIWKFPDEGACQPCPINCTHSCVDLDDKGC----- 648
Db 581 SHFKDGNCEVCKCPDGLQGANSF--IFKYADPDRECHPCNCTGCGNGPTSHDCHIYYPW 638
Qy 649 -----PAEQBASPLTSIVSAVV--GILLVVVLGVVFGILIKRQOKTRKYTMRLLQETE 701
Db 639 TGHSTLPQAHAR--TFL--IAAGVIGGLFVLVGLTFVAVYVRKSGIK-KGRALRFL-ETE 693
Qy 702 LVEPLTPSGAMPNOAMRILKETELRKVKVLGSGAFGVYKGIWIPGENVKIPVAIKVL 761
Db 694 LVEPLTPSGTAPNOAQURILKETELRKVKVLGSGAFGVYKGIWIPGETVKIPVAIKIL 753
Qy 762 RENTSPKANKEILDEAYVMAGVGSYPVSRILGICLTSTVLQVLTQMPYGCILLDHVRNRRG 821
Db 754 NETTGKANVEFMEALIMASMDPHVLRLVGLVCLSPITQLVTLQMPHGCILLEYVHEKD 813
Qy 822 RLGSQDILNWCQYAKGMSYLEDLVRLVHRDLAARNVLKSNHVKITDFGLARLLDIDET 881
Db 814 NIGSOLLNVCQYAKGMVLEERLVRDLAARNVLKSNHVKITDFGLARLLEGEK 873
Qy 882 EYHADGKVPKIMWALSILRRRTHQSDVWSYGVTVWELMTEGAKPDGIPAREIPDLL 941
Db 874 EYNADGKMPKIMWALECIHYRKHQSDVWSYGVTVWELMTEGKPDGIPAREIPDLL 933
Qy 942 EKERLPQPPCTIDVYIMVWKCMIDSECRPREFRELVSERFMRDPPQFVVIQNEB-L 1000
Db 934 EKERLPQPPCTIDVYIMVWKCMIDADSPPKELAAEFMRDPPQFVVIQNEB-L 993
Qy 1001 GPASPLDSTFVRSLEDDMGDLVDAEYLVQGGFCPPDAPAGGKWHVHRSSSTRS 1060
Db 994 KLPSFNDKFFQNLDEEDLEDDMDABEYLVQ-PAFNIPPP-----IYTSRARIIDNRS 1046
Qy 1061 GGGDLTLGLPSEEEAPRS-----PLAP-SEGAGSDV 1091
Db 1047 -----EIGHSPPPAYTTPMSGNQFVYRGGFAAEGVSVYRAPTSITPEAPVAGATAEI 1101
Qy 1092 FDGLGMAKGLQSLTHDPSPLQRYSEDPVPLPS-----ETDGYVAPLTCSPQE 1144
Db 1102 FDDSCCGTLRKVPVPHVQEDSSQRYSAQTVFAPERSPRGELDESGYTMPMRDKPKQE 1161
Qy 1145 YVQPDVVRPQPPSPRGPLPAAPAGATLERAKTILSPGKGVKVDVAFGAVENPBYLT 1204
Db 1162 YLANPVE-----ENPFVSR-----KNGDLQ-----ALDNPPEYHN 1190
Qy 1205 PQGGAAPQHPHPPA-----FSPAEDNLYWD 1230
Db 1191 ASNG-----PPKAEDVNEPLVNTFANTLGAKEYLKNILSMPEKAKAFDNPYWN 1244
Qy 1231 QDPPERGA--PPSTFKGTPT-----AENPEYL 1255
Db 1245 HSLPPRSTLQHPDYLQYSTKYFKYKQNGRIRPIVAENPEYL 1285

RESULT 7
ID_ERB4_RAT STANDARD; PRT: 1308 AA.
AC Q62956; Q622N7;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112).
GN ERB4 OR TYRO-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=98221155; PubMed=9553078;
RA Zhao Y.-Y., Sawyer D.R., Balliga R.R., Opel D.J., Han X.,
RA Marchionni M.A., Kelly R.A.;
RT "Neuregulins promotes survival and growth of cardiac myocytes.
RT Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT ventricular myocytes.";
RL J. Biol. Chem. 273:10261-10269 (1998).
RN [2]
RP SEQUENCE OF 848-901 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=91222560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system.";
RL Neuron 6:691-704 (1991).
RN [3]
RP SEQUENCE OF 1031-1198 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Frohnert P.W., Kim S.S., Corbett J.A.;
RT "Expression of neuregulins and their putative receptors, ErbB2 and
RT ErbB3, is induced during Wallerian degeneration.";
RL J. Neurosci. 17:1642-1659 (1997).
CC -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS. NRG-
CC 2. NRG-3. HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
CC NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC RECEPTORS (POTENTIAL). Type I membrane protein.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC HEART.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF041838; AAC03899.1; -.
CC EMBL; U52531; AAC53051.1; -.
CC HSSP; P11362; 1FGK.
CC InterPro; IPR000494; EGF_R_L domain.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR001245; Tyr_pkinase.
CC InterPro; IPR004019; YLP motif.
CC Pfam; PF00757; Furin-like; 1.
CC Pfam; PF00069; pkinase; 1.
CC Pfam; PF01030; Recep_L_domain; 2.
CC Pfam; PF02757; YLP; 2.
CC PRINTS; PR00109; TYRKINASE.
CC ProDom; PD000001; Euk_pkinase; 1.

DR	SMART; SM00261; FU: 4.	169	YODTILWKDIFHKNNQALTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVCA	Qy
DR	SMART; SM00219; TyKc; 1.	157	YADTIHQDILVRNPWFNMTLSTIGSSGGRCHKCTG-RCWGPTENHCOTLTRTVCAE	Db
DR	PROSITE; PS00107; PROTEIN KINASE ATP; 1.	229	GC-ARCKGPLPTDCCHEQCAAGCTGPKHSDCLACLFHNSHGICELHCPALVYNTDTFES	Qy
DR	PROSITE; PS00101; PROTEIN KINASE DOM; 1.	216	QCDGRCVGPVSDCCCHRECAAGCGPKDTCFACMFNFDSGACVTCQCPQTFVNPPTFQL	Db
KW	Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;	288	MNPBGRYTFGACVYACPNYLSTVDGSCVTLVPLHNQEVTAEDGTQRECEKSKPCARV	Qy
KW	Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.	276	EHNFNAKYTYGAFCKVCKPHNFV-VDSNSCRACPSKMEV-BENGIMCKPCCTDICPKA	Db
FT	SIGNAL	348	CYGLGMEHLREVRATVSANIOEFAGCKIFGSLAFIPESPDGDPASNTAPLOEQLOVFE	Qy
FT	CHAIN	334	CDGIGTGLMSAQTVDSSNIDKFINKINGNLIFLVGTGHPYNAIDAIDEKLVNFR	Db
FT	NP_BIND	408	TLEEITGYLYISAWPDSLPLSVFQNLQVIRGRILHNGAYSLTQGLGIGSWLGRSLREL	Qy
FT	BINDING	394	TVREITGFLNIQTWPPNMTDFSVFSLNLTIGRVLYSGLSLILKQGGITSLQFSLKEI	Db
FT	ACT_SITE	468	GSLGLIHHNTHLCFVHTVPWDQLFRPHQALLHTANRPEDECVGSLACHQICARGHCW	Qy
FT	DISULFID	454	SAGNIYITDINSNLCTYHTINMTTFLSTVNQRIYIRDNRAENCTAEGMVCNHLCSNDGCW	Db
FT	DISULFID	528	GPPTOCVNCQFLRGQECVEECRVLQGLPREVYVNAHCLPCHPEQOP-ONGSVTCFGPE	Qy
FT	DISULFID	514	GPDPDCLSCRRFSRGKICIESCNLYDGEPRFENGISICEVDSQCEKQMEDGLTCHGPG	Db
FT	DISULFID	587	ADQVACAHYKDPFPCVARCPGKPDLSYMPYIWKFPDEEGACQPCPINTHSCVDLDDK	Qy
FT	DISULFID	574	PDNCTCKSHFKDGNCKVCEKCPDVLQGANSP--IFKVADQDRECHPCHPNCTQCGNGPESH	Db
FT	DISULFID	647	GC-----PAQRASPLTSIVSAVV-GILLVVVLGVFGILIKRQOKIRKVTMR	Qy
FT	DISULFID	632	DCIYYPWTGHTSLPOHAR-TPL--IAAGVTGGFLLVIMALTFAVYVRRISIK-KGALDR	Db
FT	DISULFID	695	RLQETELVEPLTPSGAMPNQAORILKETELKRVKVLGSGAGTGVYKGIWIPDGENVKI	Qy
FT	DISULFID	688	RFL-ETELVEPLTPSGTAPNQAORILKETELKRVKVLGSGAGTGVYKGIWIPDGENVKI	Db
FT	DISULFID	755	PVAIKVLRNTSPKANKEILDEAYVMAGVSPVSRLLGICLTSTVOLVTQMLPYPGCLLD	Qy
FT	DISULFID	747	PVAIKILNETTGPKANVEFMDEALIMASVDHPHLVRLLVCLSPITQIOLVTQMLPHGCLLE	Db
FT	DISULFID	815	HVRENRLGSLQDILLNWCQIAKMSYLEVRLVHRDLAARNVLKSPNPKVITDFGLAR	Qy
FT	DISULFID	807	YVHEKDNIGSGLLNWCQIAKMSYLEVRLVHRDLAARNVLKSPNPKVITDFGLAR	Db
FT	DISULFID	875	LLDIDETEHADGKVPKIMMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIA	Qy
FT	DISULFID	867	LLEGDEKEYNADGKVPKIMMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIA	Db
FT	DISULFID	935	REIPDLLEKGERLPQPPICITIDVYIMVIMKWMIDSECRPRFRELVSFSSMARDPQRFVY	Qy
FT	DISULFID	927	REIPDLLEKGERLPQPPICITIDVYIMVIMKWMIDSECRPRFRELVSFSSMARDPQRFVY	Db
FT	DISULFID	995	IQNED-LGPASPLDSTFYRSLLDEDDMDGLVDAAEVLVPOQGFCDPP-----1041	Qy
FT	DISULFID	987	IQGDDRMKLPSPNDKSPFQNLDEEDLMDMAEYLVP-QAFNIPPIYTSRTRIDSNR	Db
FT	DISULFID	1042	-----APGAGGMVHRRSSSTRSGGDLTLGLEPSEBEEAPRSLAPSEAGSD	Qy
FT	DISULFID	1046	SETGHSPPPAYTMSGQFVYQGGFATQGG---MMPYTTATTSTIPEAPVA--QGATAE	Db
FT	DISULFID	1091	VFDGLDGMGAAGLQSLPHDPSPLQRYSEDPVPLPS-----ETDGYVAPUTCSQPQ	Qy
FT	DISULFID	1101	MEFDSGCCNGTLRKPVPVHVQEDSSSTORYSADPTVFAPERNAELDEEGYMTMHDKPKQ	Db
FT	DISULFID	1144	EYVNPQDVRPQPSREGPLPAARPAAGATLERAKTILSPGKGVVQKVFAGGAVENPEYL	Qy
FT	DISULFID	1161	EYLNPEV-----ENPFVSR-----KNGDLQ-----ALDNPPEYH	Db
FT	DISULFID	1204	TPQGAAPQPHPPA-----FSPAPDNLYW	Qy

Query Match 42.5%; Score 2911; DB 1; Length 1308;
 Best Local Similarity 44.3%; Pred. No. 8.9e-148;
 Matches 604; Conservative 190; Mismatches 384; Indels 184; Gaps 30;

1 MELA-ALCRWGLL--ALLPFGASTQVCTDMKLRPLPASPETHLMDLRHLQYCCQVQ 57
 1 MKLATGLVWGLSLVAARTVQPSASVCAGTENKLSLSLEQYRALRYENCEVYM 60

58 GNLELTLYPTNASLSFLQDIEV-----QFNFTVSWLVRPKVSAHSLRQLRIV 108
 61 GNLEITSEHRDLSFLRSIEVYGYVLVNLQRYLPL-----ENLRIL 105

109 RGTOLFEDNYALAVLDNGDPLNNTPTTVGASPGGLRELQRLSLTEILKGGVLIORNPOLC 168
 106 RGTKLVEDRYALAIPLNRYKDNF-----GLQELGLKNLTEILGGVYVDQNKPLC 156

F;658-680/Domain: transmembrane #status predicted <TM>
F;723-988/Domain: protein kinase homology <KIN>
F;731-739/Region: protein kinase ATP-binding motif
F;71,191,263,535,576,634/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;691/Binding site: phosphate (Thr) (covalent) #status predicted
F;758/Active site: Lys #status predicted
F;882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 86.2%; Score 5860; DB 1; Length 1260;
Best Local Similarity 86.0%; Pred. No. 3.3e-234;
Matches 1085; Conservative 51; Mismatches 109; Indels 16; Gaps 5;

Qy 1 MELAAALCRWGLLALLPPGAASSTVCTGDMKLRPASPTHLDMLRHLHYGCGQVQGNL 60
Db 4 MELAAALCRWGLLALLPPGIAAGTQCTGDMKLRPASPTHLDMLRHLHYGCGQVQGNL 63
Qy 61 ELTYLPTNASLSFLQDIEQVGYVLIHNAHQVQVLPQRLRVIRGTQLPEDNYALAVLDNG 120
Db 64 ELTYVPAVASLSFLQDIEQVGYVLIHNAHQVQVLPQRLRVIRGTQLPEDNYALAVLDNR 123
Qy 121 DPLNNTTPTVT-GASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILKXDIHFHKNQL 179
Db 124 DPQDNVAASPTGRTPEGLRELQRLSLTEILKGGVLIQRNPOLCYQDMVLMKDFRKNQL 183
Qy 180 ALTLIDNRSRACHPCSPMKGSRGWSSSDQSLTRTVCAAGCARCKGLPTDCCHEQ 239
Db 184 APVDIDNRSRACHPCAPACKDNHCWGESPEDCQLTGTITSCGACRCKGLPTDCCHEQ 243
Qy 240 CAAGCTGPKHSDCLACLFHNSHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCYTAC 299
Db 244 CAAGCTGPKHSDCLACLFHNSHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCYTTC 303
Qy 300 PYNLSTDVSGCTLVCPHNEQVTAEDGTQRCERKSPCARVCYGLGQXIKANSKFTGI 359
Db 304 PYNLSTEVSGCTLVCPHNEQVTAEDGTQRCERKSPCARVCYGLGMEHLRGARITSD 363
Qy 360 TELEFACKKIFGSLAFLPESFDGDPASNTAPLOEQLOVETLEETGYLIYISAWPDSL 419
Db 364 NVQBFDFCKKIFGSLAFLPESFDGDPSSGIAPLRPEQLOVETLEETGYLIYISAWPDSL 423
Qy 420 PDLVSFONLQVIRGRIHNGAYSITLQGLGISWGLSLRELGSGLALIHNNHLCFVHT 479
Db 424 RDLVSFONLRIIRGRIHNGAYSITLQGLGISWGLSLRELGSGLALIHNNHLCFVHT 483
Qy 480 VPMQDLFRNPHQALLHTANRDEDE-CVGEGLACHQLCARGHCWPGPTQCVNCSQFLRGQ 538
Db 484 VPMQDLFRNPHQALLHSGNRPEDLCVSSGLVNCNLCAHGHGWCWPGPTQCVNCSHFLRGQ 543
Qy 539 ECVBECEVRLQGLPREYVNAHCLCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVA 598
Db 544 ECVBECEVRLQGLPREYVSDKRLCHPECPQNGSVTCFGEADQCVACAHYKDSSCVA 603
Qy 599 RCPGSGVPLSYMPIWKFPPDEEGACQPCINCTHSCVDLDDKGPASORASPLTSIOYIK 658
Db 604 RCPGSGVPLSYMPIWKFPPDEEGICQPCINCTHSCVDLDDKGPASORASPLTV--FII 660
Qy 659 ANSK-----FIGITELKRQKIRKYTMRLLOETELVEPLTPSGAMPNQAWRI 708
Db 661 ATVEGVLLFLLVWVGIL-IKRRQKIRKYTMRLLOETELVEPLTPSGAMPNQAWRI 719
Qy 709 LKETELRKVKLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVM 768
Db 720 LKETELRKVKLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVM 779
Qy 769 AGVSPYVSRLLGLCLSTVQLVTLMPYXGCLLDHVRNRRGLSQDILLNWCQIAKMS 828
Db 780 AGVSPYVSRLLGLCLSTVQLVTLMPYXGCLLDHVRNRRGLSQDILLNWCQIAKMS 839
Qy 829 YLEVDRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMWALESI 888
Db 840 YLEVDRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMWALESI 899
Qy 889 LRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMI 948

Db 900 LRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMI 959
Qy 949 MVKCMWIDSECRFRRELVESEFRMARDPQRFVVIQNEIDLGPASPLDSTFVRSLEDDDM 1008
Db 960 MVKCMWIDSECRFRRELVESEFRMARDPQRFVVIQNEIDLGPSPMDSTFVRSLEDDDM 1019
Qy 1009 GDVLDAEYILVPOGFCPPDPAFGAGGMVHRRSSSTRSGGDLTLGLPSPSEEARSP 1068
Db 1020 GDVLDAEYILVPOGFCPPDPTGTSTAHRRHRSSTRSGGDLTLGLPSPSEEARSP 1079
Qy 1069 LAPSEGAGSVDFGDLGMAKGLQSLPHTDPSPLOQRYSEDPTVPLPSETDGYVAPLTCS 1128
Db 1080 LAPSEGAGSVDFGDLGMAKGLQSLPHTDPSPLOQRYSEDPTVPLPSETDGYVAPLTCS 1139
Qy 1129 PQEYVNPQVDRPQPSREGPLPAARPAAGATLERAKTLSPGKNGVVKDFAFGAVENP 1188
Db 1140 PQEYVNPQVDRPQPSREGPLPAARPAAGATLERAKTLSPGKNGVVKDFAFGAVENP 1199
Qy 1189 EYLTPOGGAAPQHPPPAFSPAFDNLVYWDQDPERGAPSTFKGTPTAENPEYLGLDVP 1248
Db 1200 EYLVPRGTASPHPSPAFSPAFDNLVYWDQDPERGAPSTFKGTPTAENPEYLGLDVP 1259
Qy 1249 V 1249
Db 1260 V 1260

RESULT 3
I48161
p-185 precursor - golden hamster
C;Species: Mesocricetus auratus (golden hamster)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999
C;Accession: I48161
R;Nakamura, T.; Ushijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishikawa, Gene 140, 251-255, 1994
A;Title: Cloning and activation of the Syrian hamster neu proto-oncogene.
A;Reference number: I48161; MUID:9413007; PMID:7908275
A;Accession: I48161
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-1254 <RES>
A;Cross-references: GB:D16295; NID:g493236; PIDN:BAA03801.1; PID:g747595
C;Genetics:
C;Gene: neu
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP
C;Keywords: domain: protein kinase homology <KIN>
F;718-983/Region: protein kinase ATP-binding motif
F;726-734/Region: protein kinase ATP-binding motif

Query Match 85.8%; Score 5833.5; DB 2; Length 1254;
Best Local Similarity 85.3%; Pred. No. 4e-233;
Matches 1075; Conservative 60; Mismatches 108; Indels 17; Gaps 3;

Qy 1 MELAAALCRWGLLALLPPGAASSTVCTGDMKLRPASPTHLDMLRHLHYGCGQVQGNL 60
Db 1 MELAAALCRWGLLALLPPGASGTVCTGDMKLRPASPTHLDIVRHLHYGCGQVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIEQVGYVLIHNAHQVQVLPQRLRVIRGTQLPEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIEQVGYVLIHNAHQVQVLPQRLRVIRGTQLPEDNYALAVLDNR 120
Qy 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILKXDIHFHKNQLA 180
Db 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILKXDIHFHKNQLA 180
Qy 181 LTLIDNRSRACHPCSPMKGSRGWSSSDQSLTRTVCAAGCARCKGLPTDCCHEQ 240
Db 181 PVDIDNRSRACHPCAPACKDNHCWGESPEDCQLTGTITSCGACRCKGLPTDCCHEQ 240
Qy 241 AAGCTGPKHSDCLACLFHNSHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCYTAC 300
Db 241 AAGCTGPKHSDCLACLFHNSHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCYTTC 300

Nature 309, 270-273, 1984
 A>Title: ATP-stimulated interaction between epidermal growth factor receptor and superoxide dismutase
 A;Reference number: A38033; MUID:84191554; PMID:6325948
 A;Contents: annotation; receptor activity
 R;Chen, W.S.; Lazar, C.S.; Lund, K.A.; Welsh, J.B.; Chang, C.P.; Walton, G.M.; Der, C.J.; Cell 59, 33-43, 1989
 A>Title: Functional independence of the epidermal growth factor receptor from a domain in the extracellular region
 A;Reference number: A33331; MUID:90003233; PMID:2790960
 A;Contents: annotation; internalization signal
 C;Comment: Binding of EGF to the receptor leads to internalization of the EGF-receptor complex
 C;Genetics:
 A;Gene: GDB:EGFR
 A;Cross-references: GDB:120610; OMIM:131550
 A;Map position: 7p12.3-7p12.1
 C;Superfamily: epidermal growth factor receptor; protein kinase homology
 C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosphotyrosine
 F;1-24/Domain: signal sequence #status predicted <SIG>
 F;25-1210/Product: EGF receptor #status predicted <NAT>
 F;25-645/Domain: extracellular #status predicted <EXT>
 F;75-300/Domain: EGF receptor extracellular domain repeat <EE1>
 F;390-600/Domain: EGF receptor extracellular domain repeat <EE2>
 F;646-668/Domain: transmembrane #status predicted <TM>
 F;669-1210/Domain: intracellular #status predicted <INT>
 F;710-975/Domain: protein kinase homology <KIN>
 F;718-726/Region: protein kinase ATP-binding motif
 F;999-1046/Region: coated-pit mediated internalization signal
 F;1047-1210/Region: inhibitory
 F;128,175,352,413,444,528,603/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F;745/Active site: Lys #status experimental

Query Match 45.8%; Score 3115; DB 1; Length 1210;
 Best Local Similarity 49.5%; Pred. No. 3.2e-121;
 Matches 624; Conservative 174; Mismatches 359; Indels 104; Gaps 21;

```

Qy 11 LLLALLPGAA--STVCTGTDMLRLPASPTHLDMLRLHYQCQVVOGNELEYPTN 68
Db 14 LLAALCPASRALEKKVCQGSNKLTLQGTEDHFLSLQRMFNCEVVLGNELEYQVN 73
Qy 69 ASLSFLDIOIRVQGVLIHNOVRQVPLRLIRIVRGTLQPDNVALAVLDNGDPLNNTTP 128
Db 74 YDLSFLKTIQEVAGVLIATNVTREIPLENLIQIRGNMYENSVALAVLSND----- 126
Qy 129 VTGASPGGLRELQRLSLEILKGGVLIQORNQLCYQDTILMKDIFHRKNQALFLDITNR 188
Db 127 ---ANKTKLKLPMRNLEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDFQNH 183
Qy 189 SRACHPCSPMKSGSRCHGESSEDCSLTRTVACAGCA-RCKGPIPTDCHHCQAGCTGP 247
Db 184 LGSQCKDPSCPNGSCWAGAGENCQKLTIIICAOQSGRCGRKSPDCCHNQCAAGCTGP 243
Qy 248 KHSCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPVNYLSTD 307
Db 244 RESCLVCRKFRDRTATCKDTPPLMLNPTTYQMDVNPGEKYSFGATCVKCPRNYVYTD 303
Qy 308 VGSCTLVPLHNOVTAEDGTQRCEKSKPCARVCYGLGMQYIKANSKFIGITILE-FAG 366
Db 304 HGSCVRACGADSYEM-EEDGVKCKCKEGPCRVKNGIGIGEFK-DSLSINATNKHFKN 361
Qy 367 CKKIFGSLAFIPESFDGDPASNTAPLOEQLOVETLEETGYLYISAWPDSLPLSVFQ 426
Db 362 CTSISGDLHLPVAFRGDSFTHTPPLDPQELDILKTVEITGFLLIQAWPENRTDLHAFE 421
Qy 427 NLOVIRGRILHNGAYSLLTQGLGSLWGLSLRSLRGLSLALIHHTHLCFVHTVPWOLF 486
Db 422 NLEIRTKOQGFSLAVSLNLTSLGLSLKLSISDGDVLTISGNKVLVANTINWKKLF 481
Qy 487 RNPQALLHTANRPEDCEVGLACHQICARGHCWGPGPTQCVNCSQFLRGQECVECRV 546
Db 482 GTSCQKTKIISNRGENSKATGVCHALCSPGCGWGPEDCVSCRNVSRGRCVCKKL 541
Qy 547 LQGLPREYVNRHCLPCHPECPQNGSVTCGPEADQCACAHYKDPFCVACPGVKP 606

```

Db 542 LESEPREFVENSECICHPPECLFOAMNITCTGRPDNCOICAHYIDGPHCVKTCPCAGVMG 601
 Qy 607 DLSYMPITWKFPEDEGACQCPINCTHSCVDLDKGCPCAEORASPLTSIQIVKANSKF--- 663
 Db 602 ENNTL-VWKYADAGHVCHLCHPNCYCTCTGPGEGCTNGPKTPSIATGMVGLLLLVV 660
 Qy 664 -IGITELKRROOKIRKYVMRLLOETELVPLTPSGAMPNOAQMRLKETELRKVKVGLS 722
 Db 661 ALGIGLFMRRHVVRKTRLLQERLEVLPLTPSGEAPNQALLRIKETEFKKIKVLS 720
 Qy 723 GAGTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMAGVSPYVRLGI 782
 Db 721 GAGTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMAGVSPYVRLGI 780
 Qy 783 CLTSTVOLQMLPYGCLLDHVRNRLGSLQDLNMCQIAKGMYSLEDVRLVHRDLAA 842
 Db 781 CLTSTVOLQMLPYGCLLDHVRNRLGSLQDLNMCQIAKGMYSLEDVRLVHRDLAA 840
 Qy 843 RNVLVKSPNHVKITDFGLARLLDIDETEHADGGKVPDKWMALESILRRRFTHQSDVMSY 902
 Db 841 RNVLVKTPQHKITDFGLARLLDIDETEHADGGKVPDKWMALESILRRRFTHQSDVMSY 900
 Qy 903 GVTVMELMTCAKPYDGIIPAREIPDLLEKGERLPDPDICTIDVYVMVKWMDSECRPR 962
 Db 901 GVTVMELMTCAKPYDGIIPAREIPDLLEKGERLPDPDICTIDVYVMVKWMDSECRPR 960
 Qy 963 FRELVSFERNARDPQRFVVIQ-NEDLGASPLDSTFYRSLLEDMDGLVDABEYLVPQ 1021
 Db 961 FRELVSFERNARDPQRFVVIQ-NEDLGASPLDSTFYRSLLEDMDGLVDABEYLVPQ 1020
 Qy 1022 QGFPCPDPAFCAGMWHHRSSSTRSGGDLTLGLEPSEEAAPRSPAPSEAGSDVFD 1081
 Db 1021 QGFF-----SSPSTSRTELLSSLSATSN--N 1044
 Qy 1082 GDLMGNAKGLQSLPHDPSLQRYSEDPTVPLPSET--DGYVAPLTCSPQEVNPDV 1139
 Db 1045 STVACIDRNLGSCPTKEDSFQRYSDPTGALTEDSIDDTFL-----VPVEYINQ-SV 1097
 Qy 1140 RPOPPSPREGPLPAAPAGATLERAKTSLGPKGVKDVFAFGAVENPVL-TPQGGAA 1198
 Db 1098 PKPAGSVQNVVHQNPLN-----APSRDPHQD--PHSTAVGNPEYLVNQ---- 1143
 Qy 1199 PQHPPPAFSPAFDNLYWDQ-----DP-----PERGAPPSTFKGTPTAENPEY 1242
 Db 1144 ----PTCVNSTFSDPAHWAQKSHQISLDNPDYQDFFPKKAKPNCIFKGS-TAENAEY 1197
 Qy 1243 L 1243
 Db 1198 L 1198

RESULT 5
 A53183
 epidermal growth factor receptor precursor - mouse
 C;Species: Mus musculus (house mouse)
 C;Date: 06-Jan-1995 #sequence revision 06-Jan-1995 #text change 18-Jun-1999
 C;Accession: A53183; A43818; S24942; A28941; S45325; I49543
 R;Luetteke, N.C.; Phillips, H.K.; Qiu, T.H.; Copeland, N.G.; Earp, H.S.; Jenkins, N.A.;
 Genes Dev. 8, 399-413, 1994
 A;Title: The mouse waved-2 phenotype results from a point mutation in the EGF receptor
 A;Reference number: A53183; MUID:94170986; PMID:8125255
 A;Accession: A53183
 A;Molecule type: mRNA
 A;Residues: 1-1210 <LUE>
 A;Cross-references: GB:003425
 R;Avivi, A.; Lax, I.; Ullrich, A.; Schlessinger, J.; Givol, D.; Morse, B.
 Oncogene 6, 673-676, 1991
 A;Title: Comparison of EGF receptor sequences as a guide to study the ligand binding site
 A;Reference number: A43818; MUID:91232866; PMID:2030916
 A;Accession: A43818
 A;Molecule type: mRNA
 A;Residues: 1-714 <AVI>
 A;Cross-references: GB:X59698

R;Eisinger, D.P.; Serrero, G.
 submitted to the EMBL Data Library, June 1992
 A;Reference number: S24942
 A;Accession: S24942
 A;Molecule type: mRNA
 A;Residues: 969-971,'K', 973-1115,'D' <EIS>
 A;Cross-references: EMBL:Z12608
 R;Heisermann, G.J.; Gill, G.N.
 J. Biol. Chem. 263, 13152-13158, 1988
 A;Title: Epidermal growth factor receptor threonine and serine residues phosphorylated in
 A;Reference number: A28941; MUID:88330814; PMID:3138233
 A;Accession: A28941
 A;Molecule type: protein
 A;Residues: 689-694,'X', 696-704,'L', 706-707;989-992,'XX', 995-996,'X', 998-1000;1002-1009,
 R;Hibbs, M.L.; Dunn, A.R.; Alexander, W.S.
 submitted to the EMBL Data Library, April 1994
 A;Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor
 A;Reference number: S45325
 A;Accession: S45325
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-971,'K', 973-1210 <VER>
 A;Cross-references: EMBL:X78987; NID:9488830; PIDN:CAA55587.1; PID:9488831
 R;Paria, B.C.; Das, S.K.; Andrews, G.K.; Dey, S.K.
 Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993
 A;Title: Expression of the epidermal growth factor receptor gene is regulated in mouse b
 A;Reference number: 149643; MUID:93128380; PMID:7678348
 A;Accession: 149643
 A;Status: translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 12-20,22-132 <RES>
 A;Cross-references: GB:L06864; NID:g193001; PIDN:AAA53029.1; PID:g567201
 C;Genetics:
 A;Gene: EGFR
 C;Superfamily: epidermal growth factor receptor; protein kinase homology
 C;Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphop
 F;1-24/Domain: signal sequence #status predicted <SIG>
 F;648-670/Domain: transmembrane #status predicted <TM>
 F;712-977/Domain: protein kinase homology <KIN>
 F;720-728/Region: protein kinase ATP-binding motif
 F;680,695/Binding site: phosphate (Thr) (covalent) #status experimental
 F;697,1070,1071/Binding site: phosphate (Ser) (covalent) #status experimental
 F;993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental
 F;1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental
 F;1157/Binding site: phosphate (Tyr) (covalent) #status experimental

Query Match 45.4%; Score 3082; DB 2; Length 1210;
 Best Local Similarity 49.5%; Pred. No. 7.3e-120;
 Matches 630; Conservative 164; Mismatches 362; Indels 116; Gaps 25;

Qy	11	LLLLLPGAA--STQVCTGDMKRLPASPEHLDMLRLHYQGVQVQGNLELTILPTN	68
Db	14	LLTALCAAGGALEKKVCGTSNRLTQLGTGFEDHFLSLQRYNNCEVVLGNLEITYQVN	73
Qy	69	ASLSFLQDIQEVQGVLLTAHNOVRQVPLQRLIRVGTOLFFEDNVALAVLDNGDPLNNTTP	128
Db	74	YDLSEFLKTIQEVAGVLLALNTVERIPLENQITRGNALYENTYALALSN-----	124
Qy	129	VTGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDI-----FHKNNQLALFLI	184
Db	125	-YGNRTGLRELPMRELQELLIGAVRFNSNPILCNMDTIQWRDIQVNVFMSNMSMDL---	180
Qy	185	DTNRSRACHPCSPCKGSRGWESSEDCQSLTRTVACGGCA-RCKGPLPTCCHEQCAAG	243
Db	181	-QSHPSKCPKCDPSPNGSCWGGGEENCOKLTKICAOQCCHRCGRSPSCCHNQCAAG	239
Qy	244	CTGPKHSDCLACLFHNHSGICELHCPALVLTNTDTFESMPNPEGRYTFGASCVTACPNY	303
Db	240	CTGPRESDCLVCKQFQDEATKDTCPPLMLNPTTYQMDVNPESKYSGATCVKCKPNY	299
Qy	304	LSTDVGSSTLVCLPLHNQBVTAEDGTQRCCKSKPCARCYGLGMQYIKANSKFIGITELE	363
Db	300	VVTDHSGSVACGPDYEV-BEDGIRKCKKCDGPKCKVCGNIGIGIEFK-DTLISINATNIK	357

364 -EAGCKKIFGSLAFLPESFDGDPASNTAPLQBPOLQVFEETLEETTYLYIISAWPDSLPLD 422
 358 HFKYCTAISGDHLHLPVAFKGDSTFTRPPLDPRELEILKTVKEITGSLQLQAMPDNWTDL 417
 423 SVFQNLQVIRGRILHNGASLTLLQGLISWLGSLRSLRELGSGLALIHNNHLCFVHTVPW 482
 418 HAFENLEIIRGRTHQGFSLAVVGLNITSLGLRSLKEISDGDVLIISGNRLCVANTINW 477
 483 DQLFNPHQALLHTANPEDECVEGEGLAHQLCARGHCWGPGTQCVNCSQFLRGQSCVE 542
 478 KKLFGTNPQKTKIMNNAEKDKAVNVHVCNPLCSSEGCWGPEDRDCVSCQNVSRGECVE 537
 543 ECVLQGLPREYNARHCLCHPECOFGNSVTCFGEADQCVACAHYKDPFFCVARCP 602
 538 KCNILEGEPREFVENSECIOCHPECLQAMNICTGRPCDNCIQCAHYIDGPHCVKTCPA 597
 603 GVKPDLSYMPIWKFPFDEEGACQPCPCINCTHSCVDLDDKGCPCAPQASPLTSIQY- IKANS 661
 598 GINGENNTL-VMKYADANNVCHLCHANCTYGCAGPGLQGCVEWPSGPKIPSIATGIVGGL 656
 662 KFI-----GITELKRRQOKIRKVTMRLLQETELVEPLTFTSGAMPNQAOHRLKETELRK 716
 657 LFTVVALGIGLFMRERRHIVKRTRELLQERELVEPLTFTSGEAPNQAHRLKETETBFKK 716
 717 VKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVGSYV 776
 717 IKVLGSGAGFTVYKGLWIPGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPHV 776
 777 SRLIGICLTSTVQLVLTQMPYGCILLDHVRENRRGLSGQDLLNMCQIAKMSYLEDVRLV 836
 777 CRLLIGICLTSTVQLITQMPYGCILLDVREHKONIGSQYLLNMCVQIAKGNVLEDRRLV 836
 837 HRDLAARNVLKSPNVKVIITDFGLARLLDIDETEHADGCKVPIKWMALLESILARRRTHQ 896
 837 HRDLAARNVLKTPQHVKIITDFGLAKULGAEEKEYHAEGGKVPKWMALLESILHRIYTHQ 896
 897 SDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITDVYMIWVKWMD 956
 897 SDVMSYGVTVWELMTFGSKPYDGIIPASDISILEKGERLPQPPICITDVYMIWVKWMD 956
 957 SECRPRELVSFBSRWARDQRFVVIQ-NEDLGASPLSDSTFYRSLLEDMDMDGLVDAAE 1015
 957 ADSRKPREFILBFSQWARDQRFVVIQGDHMLPSPTDSNFYRALMDEEDMEDVVDAD 1016
 1016 EYLVPOQGFCDPAPCAGGWHHRSSSTRSGGDLTLGLEPSEEAARSPPLASEGA 1075
 1017 EYLIPQOGFF-----NSPST-----SRTPLLSLSA 1042
 1076 GSDVFDGDLGMAKGLQSLPTHDPSPQRYSEDPVLPSET--DGVAPLTCSPQPEY 1133
 1043 TSN-----NSTVACINRNGSCRVKEDAFQRYSDPTGAVTEDNIDDAFL-----PVPEY 1092
 1134 VNQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVYKDVAFGGAVENPEYL-T 1192
 1093 VNQ-SVPEKPRAGSVQNVFVHNQPLHP-----APGRDLHYQN--PHSNAVGNPEVLT 1141
 1193 PQGAAPQPHPPAFSPAFDNLYWDO-----DP-----PERGAPPSTFKGTPT 1236
 1142 AQ-----PTCLSSGFSNPAWIKGSHQMSLDNPDYQODFPFKETKPNGIFKG-PT 1191
 1237 AENPEYLGLDVP 1248
 1192 AENAEYLRVAPP 1203

RESULT 6
 TVCHLV
 epidermal growth factor receptor precursor - chicken
 N;Contains: protein-tyrosine kinase (EC 2.7.1.12) erbb
 C;Species: Gallus gallus (chicken)
 C;Date: 28-Feb-1986 #sequence_revision 05-May-1995 #text_change 04-Feb-2000
 C;Accession: A27720; A00643

C;Keywords: ATP; growth factor receptor
F;716-981/Domain: protein kinase homology <KIN>
F;724-732/Region: protein kinase ATP-binding motif

Query Match 43.0%; Score 2925.5; DB 2; Length 1308;
Best Local Similarity 45.2%; Pred. No. 2.2e-113;
Matches 608; Conservative 175; Mismatches 385; Indels 177; Gaps 28;

```
Qy 9 WGLLLALLPPGAA-----STQVCTGTDMLRLPASPEHLDMLRHLVQCQVVOGNIETY 64
Db 8 WWSVLLVAAGTVQPSDSQSCVAGTENKLSLSLDEQQYRALRYKYENCEVVMGNLEITS 67
Qy 65 LPTNASLSFDIOIQVQGVYLIAHNQVQVPLQRLRIRVGTQFLFEDNYALAVLQNGDPLN 124
Db 68 IEHNRDLSFVSREVTGVYVALNQFRLPLENLRIIRGTLYEDRVALAIFLNYKDG 127
Qy 125 NTPVTGASPGRLRLQLRSITELKGVLIQORNQPCYQDTILWKDIFHKNNQALATLI 184
Db 128 NF-----GLQELGLKNLITEILNGVYVDQNKFLCYADTIHQDIVRNPWPSNLTIV 178
Qy 185 DTRSRACHPCSPMKGSRGSESDCOSLTRTVTCAGGC-ARKGPLPTDCHEOCNAG 243
Db 179 STNGSSGGRCHKCTG-RCWGPTENHCOTLTRTVCAEQDGRCYGYPVSDCCCHREAGG 237
Qy 244 CTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPNY 303
Db 238 CSGPKDTCFACMFNDSGACVTCQPTFVYNPTTFQLEHFNNAKYTFAGFCVKKCPNF 297
Qy 304 LSTDVGSCTLVCPHLNQEVTAEDTQRCCEKSKPCARCYCGLGMOYIKANSKFIGITELE 363
Db 298 V-VDSSSCVRACPSKMEV-BENGIMCKPCTDICKPKADGIGTGLMSAQTVSSNIDK 355
Qy 364 FAGCKIIGSLAFLPESFDGPASNTAPLOEQLOVPETLEITCYLYISWPSLPLDS 423
Db 356 FINCTKINGNLIFTVGIHGPYPYNAIEADPEKLVNFTREIVTEIFGLNIQSWPNMTDFS 415
Qy 424 VFQNLQVIRGRIHNGAYSLTLQIGISWGLRSLRELGSGLALHHTHLCFVHTVPWD 483
Db 416 VFSNLVTIGRVLVYSGLSLLILKQOITSLOFQSLKEISAGNIYITDNSLCYHTINWT 475
Qy 484 QLFRNPHQALLHTANRDECVGEGLACHQICARGHCWGPQTCVNCQSQFLRGOECVEE 543
Db 476 TLFSTINQIRIVTRDNRKAENCTAEGMVCNHLCSGDCWGPQDQCLSRFRSRRGICIES 535
Qy 544 CRVLQGLPREYVNAHCLPCHPEQOP-QNGSVTCGPEADOCVACHYKDPFCVARCP 602
Db 536 CNLYDGEPRFENGISICEVCDPQCEKMDGLLTHCGPDPDNCCKSHFKDGNPCVEKCPD 595
Qy 603 GKVPDLSTYMPIWKFPDEEGACQPCPINCTHSCVLDDKGC-----PAEORASPL 651
Db 596 GLOGANSF--TFKYADPDRCHPCNPCTQCGNGFTSHDCIYYPWTGHTLPOHAR-TPL 652
Qy 652 TSIQVIKANSKFI-----GIT--ELKRROOKIRKYTMRELLQETELVELPTPSGAMPNOAQ 705
Db 653 TAAAGVI--GGLFILVIGLTFEAVYVRRKSIKKRALRRFL-ETELVELPTPSGTAPNOAQ 709
Qy 706 MRILKETELRKVKVLGSGAFCTYKGIWIPDGENVKIPVAIKVLENTSPKANKEILDEA 765
Db 710 LRLIKETELRKVKVLGSGAFCTYKGIWIPDGETVKIPVAIKVLENTSPKANKEILDEA 769
Qy 766 YVMAGVGSPPYVSRLLIGICLTSTVQLVTLQMPYGCILLDHRVNRGRGLGSQDLLNWCMIQAK 825
Db 770 LIMASMDHPHLVRLLVGLVCLSPITQLVTLQMPHGCLLEYVHEKONIGSQLLNWCQVIQAK 829
Qy 826 GMSYLEDLVRLVHRDLAARNVLKSPNNHKITDFGLARLLDIDETEHADGKGVPIKMMAL 885
Db 830 GMYLEERLVRDLAARNVLKSPNNHKITDFGLARLLDIDETEHADGKGVPIKMMAL 889
Qy 886 ESILRRRTHOSDVMYSYGVTWELMTFGAKYVDGIPAREIPDLLEKGERLPQPTCTIDV 945
Db 890 ECIVHRKTHOSDVMYSYGVTWELMTFGKYPDGIPTREIPDLLEKGERLPQPTCTIDV 949
Qy 946 YIMVVKCWMIDSECRPRELVSFSESRMARDPQRFVVIQNE-D-LCPASPLDSTFVRSILLE 1004
```

```
Db 950 YIMVVKCWMIDADSRPKFKELAAEFSESRMARDPQRYLVIQDDRMKLPSPNDSKFFQNLLD 1009
Qy 1005 DDMGDLVDAAEEVYLPOQGFCDPPAPGAGVMVHRRSSSTRSGGDLTLGLPSPSEEA 1064
Db 1010 EEDLEDMDAAEVLVP-QAFNIPPP-----IVTSRARIDSNRS-----EIGHSPPPAYT 1057
Qy 1065 PRS-----PLAP-SEGAGSDVDFDGLGNGAAKGLQSL 1095
Db 1058 PMSGNQFYVRDGGFAABQGVSVPYRAPSTIPEAPVAAQTALIFDDSCCNGTLRKVPAP 1117
Qy 1096 PTHDPSFLORYSESDPTVPLFS-----ETDGVYVAPLTCSPQPEYVNPQDVRPPOPPSPRE 1148
Db 1118 HVQSDSSTQYSADPTVFAPERSPRGELDEEGYMTWRDKPKQEYLNPE-----E 1168
Qy 1149 GPLPAARPAGATLERAKTILSPGKNGVVKQVFAFGGAIVENPEYLTPOGGAAPQHPPPA-- 1206
Db 1169 NPFVSR-----KNGDLQ-----ALDNPEYHNASNG-----PPKAED 1200
Qy 1207 -----FSPAFDNLVYWDODDPPERGA--PPSTFK 1232
Db 1201 EYVNEPLYLNTFANTLGAAYLKNILMSPEKAKKAFDNPDYWNHSLPPRSTLQHPDYLQ 1260
Qy 1233 GTPT-----AENPEYL 1243
Db 1261 EYSTKYFYKQNGRIRPIVAENPEYL 1285

RESULT 8
S06142
protein-tyrosine kinase (EC 2.7.1.112) mkp-y precursor - southern platyfish
N;Alternate names: epidermal growth factor receptor homolog; kinase-related transforming
C;Species: Xiphophorus maculatus (southern platyfish)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 18-Feb-2000
C;Accession: S06142; S13809
R;Wittbrodt, J.; Adam, D.; Malitschek, B.; Maeueller, W.; Raulf, F.; Telling, A.; Roberts
Nature 341, 415-421, 1999
A;Title: Novel putative receptor tyrosine kinase encoded by the melanoma-inducing Tu loc
A;Reference number: S06142; MUID:90015140; PMID:2797166
A;Accession: S06142
A;Molecule type: DNA
A;Residues: 1-1166 <MIT>
A;Cross-references: EMBL:X16891; NID:g65290; PIDN:CAA34770.1; PID:g65291
R;Adam, D.; Maeueller, W.; Scharlt, M.
Oncogene 6, 73-80, 1991
A;Title: Transcriptional activation of the melanoma inducing Xmrk oncogene in Xiphophoru
A;Reference number: S13807; MUID:91125882; PMID:1846957
A;Accession: S13809
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 821-1025, 'N', 1027-1098, 'A', 1100-1166 <ADA>
A;Cross-references: EMBL:X56319; NID:g65284; PIDN:CAA39763.1; PID:g65285
C;Genetics:
C;Gene: mrk
A;Map position: Y
A;Introns: 872/3; 898/1; 947/1; 979/3; 1025/3; 1056/1
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; growth factor receptor; phosphotransferase; transmembrane protein; tyro
F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-1166/Product: kinase-related transforming protein (Tu) #status predicted <MAT>
F;707-972/Domain: protein kinase homology <KIN>
F;715-723/Region: protein kinase ATP-binding motif

Query Match 38.8%; Score 2635; DB 1; Length 1166;
Best Local Similarity 45.0%; Pred. No. 1.8e-101;
Matches 570; Conservative 160; Mismatches 393; Indels 144; Gaps 30;

Qy 4 AALRCWGLLALLPPGAASL---OVCTGTDMLRLPASPEHLDMLRHLVQCQVVOGNI 59
Db 8 AALLQ--LLLVLSISRCCSTDPDRKVCQGSINQMTM---LDNHYLKKKKVYSGCNVLEN 62
Qy 60 LELTYLPTNASLFLQDIOEVQGVYLIAHNQVQVPLQRLRIRVGTQFLFEDNYALAVLDN 119
```

Db 63 LEITYTOENQDLSEFQSTQEVGGVYLIAMNEVSTIPLVNLRLIRQNLNLYEGNFLLVMSN 122
Qy 120 GDPLNNTTPVTGASPGGLRELQLRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNQL 179
Db 123 YQK-NPSSP--DVYQVGLKQLQLSNLTELSSGVKVSHPNLLCNVEINWWDIVDKTSNP 179
Qy 180 ALTLIDNRSRACHPCSPMKGRCSWGSSESDCSLRTTVCAGGC-ARCKGPLPTDCCHE 238
Db 180 TMNLI PHAFERQOCQKDRHGVNGSWAPFGHCQKFTKLLCAEQCNRRRCRGPKEIDCCNE 239
Qy 239 QCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTA 298
Db 240 HCAGCTGPRATDCLACRDFNDGTCCTKCTPPKIYDIVSHQVNDPNIKTFGAACVKE 299
Qy 299 CPYNYSLTDVGSCTLVCPHNNQEVTAEDGTQRCCKSPCARVCYVGLGMQVIKANSKF 358
Db 300 CPSNVVTE-CACVRSWSCAGMLEVD-ENGKSRPCDGVCPKVCDDIGIGSL-SNTIAVN 356
Qy 359 ITEL-EFAGCKKIFGSLAFLEPESDGDPASTAPLOPQLOVFTLEBITGYLISAWPD 417
Db 357 STNIRSFNCTKINGDIILNRSPEGDPHYKIGTMDPEHLNWLTTVKBITGYLVIMWPE 416
Qy 418 SLPDLVFNQLQVIRGRILHNGAYS-LTLOGLGISWGLRSLRELGLGLALIHNTLHCF 476
Db 417 NMTLSLVFNQLLEIIRGRTTFRGFSFVVQVRHLQWLGLRSLKEVSAGNVILKNTLQRY 476
Qy 477 VHTVPWOLFNRPHQALLHTANRDECECVGBGLACHQLCARGHGWPGPTQCVNCSOFLR 536
Db 477 ANTINWRLEFSEDSIEYDART-----ENQTCNNECSESDGCW-PGPTWCVSLHVD 528
Qy 537 GQCVCEBRCVLOGLPREVYNARHCLPHCEPQONGSVTCFPGBADOCVCAHYKDPDFC 596
Db 529 GGRCVASCNLLQGBPREAQVQDRCVQCHOECLVQDLSLTCYGPFGPANGSKSAHFQDGP 588
Qy 597 VARCPGVKPLSYMTKWPDEEGACQCPINCTHSCVDLDDKCCPAE-----QRASPL 651
Db 589 IPRCPHGLGDGTL-INKYADKMGQCCQPCQHCNCTQCGSGGLSGCRGDIVSHSLAVGL 647
Qy 652 TSIOVIKANSFIFGITELKRQKIRKYKTRMRLLQETELVEPLTPSGAMPNQAQMRILKE 711
Db 648 VSGLLITIVALLIVLLRRRIK-RKETIRCLQEKELVEPLTPSGAQNQAFLRIKE 706
Qy 712 TELRKVKVLSGAGFTVYKGIWIPDGENVKIPVAKVLRNTPSKANKEILDEAYVWAGV 771
Db 707 TEFKKDRVLGSGAGFTVYKGLWNPNGENIRIPVAKVLRNTPSKVNOEVLDEAYVWASV 766
Qy 772 GSPVYSLGLGLTSLVOLTQMPYGCCLLDHVRNRCGLSQDILLNWCQIAGMSVLE 831
Db 767 DHPVHVRLLGLTSAVOLVTLQMPYGCCLLDYVRQHOERICQWLLNWCQIAGMNYLE 826
Qy 832 DVLVLRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGQKVPKKNMALESILRR 891
Db 827 ERHLVHRDLAARNVLLKNPNHVKITDFGLSKLLTADKEYQADGQKVPKKNMALESILQ 886
Qy 892 RPTHQSDVMSYGVTVWELMTGAKPYDGPAREIPDLLEKGERLPQPPICHTIDYVMIMVK 951
Db 887 TYTHQSDVMSYGVTVWELMTGSKPYDGPAPAKETASVLENGERLPQPPICHTIEVYMIK 946
Qy 952 CWMIDSECRPFRELVEFSRWARDQPFVVIQNEIDLGPASPLDSTFFRSRLLEDGMDL 1011
Db 947 CWMIDPSRRPFRELVEFSGQARDPSRYLIQG---NLPSLSDRRFLSRLSSDD--DV 1001
Qy 1012 VDABEYLVPQGGFFCPDPAPGAGGVVHRRSSSTRSGGDLTLGLEPSEEAAPRSLAP 1071
Db 1002 VDAEYLLPYKRI-----NRQS-----EFCIP 1024
Qy 1072 SEGAGSVDFDGLGMAAGKIQSLPTHDPSPLOYSEDPV-PLPSETDGYVAPLTCSPQ 1130
Db 1025 PTGH-----PVRENSITLRNISDPTQNALEKOLDGH-----1055
Qy 1131 PEYNQDVRPOP-----PSPRE-----GPLP-AARPAGATLERAKTLPFGKNGVVKD 1177
Db 1056 -EYVNPQGETSRLSDIYNPNYEDLTDGWDGPFVSLSQEAEITNFSREYLTNTQNSL--- 1111

Qy 1178 VFAFGGAVENPEYLTPQGGAAPOHPHPAPAFDNLNLYYWDQPPPERGAPSTFTKGTPTA 1237
Db 1112 PLVSSGSMDDPDY---QAG-----YQAAF-----LPQTGALTGNMFLPAA 1149
Qy 1238 ENPEYLG 1244
Db 1150 ENLEYLG 1156
RESULT 9
A36223
kinase-related transforming protein (erbB3) (EC 2.7.1.1-) precursor - human
C:Species: Homo sapiens (man)
C>Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000
C:Accession: A36223; I59164
R:Kraus, M.H.; Isings, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.
Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989
A:Title: Isolation and characterization of ERBB3, a third member of the ERBB/epidermal g
A:Reference number: A36223; MUID:90083234; PMID:2687875
A:Accession: A36223
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1342 <KRA>
A:Cross-references: GB:M29366
R:Plowman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todaro, G.J
Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990
A:Title: Molecular cloning and expression of another epidermal growth factor receptor-re
A:Reference number: I59164; MUID:90311312; PMID:2164210
A:Accession: I59164
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-559, 'G', '561-957', 'F', '959-1063', 'G', '1065-1342 <RES>
A:Cross-references: GB:M34309; NID:g183990; PIDN:AAA35979.1; PID:g306841
C:Genetics:
A:Gene: GDB:ERBB3; HER3
A:Cross-references: GDB:I19880; OMIM:190151
A:Map position: 12q13-12q13
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog
C:Keywords: Arp; phosphotransferase
F:707-972/Domain: protein kinase homology <KIN>
F:715-723/Region: protein kinase Arp-binding motif
Query Match 35.3%; Score 2398.5; DB 2; Length 1342;
Best Local Similarity 40.4%; Pred. No. 1.1e-91;
Matches 530; Conservative 189; Mismatches 456; Indels 137; Gaps 33;
Qy 10 GLLALLPPGAA--STQVCTGTDKMLRLPASPETHLDMLRHLVYGGCOVVOGNLELTYLPT 67
Db 11 GLLFSLARGSEFVNSQAVCPGTLNGLSVTDAENQVQTLKYERCEVVMGNLEIVLTGH 70
Qy 68 NASISFLQDIQEVQGVYVLIHANNVQVPLQRLRIVRGTLQFEDNYALAVLDNGDPLANNTT 127
Db 71 NADLSFLQWIKRETVGYVLYVAMNFEFTLPLNLRVVRGTQVYDGKFAIFVM-----LNYNT 125
Qy 128 PVTGASPGGLRELQLRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNQALTLIDTN 187
Db 126 ----NSSHALRQLRLTOLTILSGSVYIEKNDKLCMDTIDMRDIVDRD---AEIVVKD 178
Qy 188 RSRACHPCSPMKGRCSWGSSESDCSLRTTVCAGGC-ARCKGPLPTDCHEQCAAGCTG 246
Db 179 NGRSCPPCHEVCKG-RCWGPGSEDCOTLTKTICAQCNHGHCFGNPNQOCCHDECAGCGSG 237
Qy 247 PKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTCAPYNYLST 306
Db 238 PQDTDFACRHFNDSGACVPRCPQLVYNKLTQLEPNHTKYQYGGVCVASCNPHFV-V 296
Qy 307 DVGSCTIVCLPHNQEVTAEDGTQRCCKSPCARVCYVGLGMQVIKANSKF--IGITEL- 363
Db 297 DQTSVCVACPDKMEVD-KNGLKWCPCGGLCPKACBGTG-----SGSRQTVDSSNIDG 350
Qy 364 FAGCKKIFGSLAFLEPESDGDPASTAPLOPQLOVFTLEBITGYLISAWPDSLPDLS 423

351	Db	FVNC	KILGNLDFLITGLNGDPWHKI	PALDPEKUNVFRTRVREITGYLNIQSWPPHMFNS	410
424	Qy	VFQNLQVIRGRI	LHNAYS-LTLOGLGSM	LGRSLRELGSGLALIHNNTHLCFVHTVPW	482
411	Db	VFSNLTTIGRSL	YNRGFSLLIMKNLVN	SLGFRSLKEISNAGRIYISANRQLCYHHSLNW	470
483	Qy	DQLFRNPQALLHTA	-NRPEDEVCEGEG	LACHQLCARGHCWGPPTQCVCNSQFLRGQECV	541
471	Db	TKVLRGTEBR	LDIKHNPRDRCD	VAEGKVCDPCLCSSGCGWPGQCLSCRNYSRGVCV	530
542	Qy	BECHVLOGLPRE	YNARHCLPCH	CEQOPNGSVTCFGEADOCVACAHYKDPFPCVACRP	601
531	Db	THCNFLNGEPR	FAHEACFSCH	PECFMEGTATCNGSGSDTCAQCAHFRDGHCVSSCP	590
602	Qy	SGVKPDLISYMP	IKAFPPDEEGAC	QPCPNCTHSCVDLDDKGC	653
591	Db	HGVUG--AKGPI	YKTPDVQNECR	FCHENCTOGCKGPELQDCLGTLVLIGKTHLTMA	648
654	Qy	IQYIKANSKF	IGITELKRQOKIR	-KYTMRELLQETELVEPLTPSGAMPQAO	712
649	Db	IAGLVVIFM	LGGTFLYWRGRI	QNKRAMRRLYERGESSIEPLDPS-EKANKVLARI	707
713	Qy	ELRKVKVLGSGA	FGTVYKGIWIP	GENVKIPVAIKVLRNTSPKANKEIID	772
708	Db	ELRKLKVLGSG	VFTVHKGVWIP	EGESIKIPVCIKVIEDKSGRGSFOAVTDHMLAIGSLD	767
773	Qy	SPVYSRLIG	ICLTSTVOLV	TOLMPYGCLLDHVRNRRGLSGODLNNCMOIAKGM	832
768	Db	HAHIVRL	LGLCPGSSQLV	TQYLPGLSLLDHVRQHRGALGQLLNNGVQIAKMYL	827
833	Qy	VRLVHRDLA	ARNVLVKS	PNHKITDIFGLARLLDDIETEHADGKVP	892
828	Db	HGMVHRNLA	ARNVLLKSPS	QOVADFGVADLLPDDKQLLYSEAKTPIK	887
893	Qy	FTHOSDVWSY	GVTVWELMT	FGAKYDGPAREIPDLEKGERLPQPICTIDV	952
888	Db	YTHOSDVWSY	GVTVWELMT	FGAEYPAGRLAEVDPDLLEKGERLAQ	947
953	Qy	WMIDSECR	PREFELVSFR	SESMARDPQRFVITQNEIDLGPA---SPLD	1009
948	Db	WMIDENIR	PTEKLANEFT	RMARDPPRLVLIKRS-GFGIAGPEPHGLTN	1006
1010	Qy	DLVDAEYLV	PQOGFFCPD	PAPGAGVMVHRRHSSTRSGGDI	1068
1007	Db	PELDLDL	DLAEED-----	NLAITLGSALS	1047
1069	Qy	LAPSEGAGS	VDVFDGLG	MAAKGLOSLPTHD-PSPLORYS	1211
1048	Db	LSPSSGY-M	PMNQNLG	ESCOESAVSGSERCP	1100
1122	Qy	VA-----	PLTCS	POPE---YVNPQDVR	1150
1101	Db	VTGSEAE	LQEKVWCR	SRSRSPRPGDSAYHSQRHSLLT	1160
1151	Qy	LPAARPA	GATLERAKT	LSP-KGNV-----KDVFAFGG	1203
1161	Db	MPDTHL	KGTPSSREG	TLSSVGLSVLTGTEEDED-----E	1211
1204	Qy	PPAFSPAF	DNLIYYWD-----	QDPPERGAPSTFKGTPTAENPEYL	1243
1212	Db	PPSSLE	BELGYEYMDV	SGDLSASLGSTQSCPLHPVIMPTAGT	1263

RESULT 10

JC4387
 Epidermal growth factor receptor homolog precursor - rat
 N:Alternate names: ErbB3 protein; HER3 protein
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998
 C:Accession: JC4387
 R:Hellayer, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.
 Gene 165, 279-284, 1995
 A:Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protein.

Db 699 ARIPEKTELKLVKLVGSGVGVTHKGIWIPGESIKIPVICIKVIEDKSGRGSFOAVTDHM 758
Qy 766 YVMAGVSPYVSRLLIGICLTSTVOLVQLTQMPYGCILLDHVRNRRGLSGQDILLNMCWQIAK 825
Db 759 LAVGSLDHAHIVRLILGLCPGSSQLQVLTQYPLUGSLLDHVKHQRETLGQPLLNGVQIAK 818
Qy 826 GMSVLELDVRLVHRDLAARNVLKSPNPKITDFGLARLLDIDETEHADGKVPKIKWAL 885
Db 819 GMYLLEHSHVHRDLALANVLKSPSQVQVADFGVADLLPPDKQLLHSEAKTPIKWWAL 878
Qy 886 ESILRRRTHQSDVMSYGVTVWELMTFCAGKPYDGIIPAREIPDLLEKGBRLPPQPICTIDV 945
Db 879 ESIHFKYTHQSDVMSYGVTVWELMTFCAGPYAGLRLAEIPDLLEKGBRLAQPQICTIDV 938
Qy 946 YIMVWKWMDSECRPRFRELVSFMRARDPQFVVIQNDLGPASPLDSTFYRSLLED 1005
Db 939 YIMVWKWMDSECRPRFRELVSFMRARDPQFVVIQNDLGPASPLDSTFYRSLLED 995
Qy 1006 DDMGDLVDAEYLVPOQGFPCDPAPGAGGMVHRHRSSTSRSGGDLTLGLEPSEE--- 1062
Db 996 KEL-----QEALELEPEL-----DLDLLEAEDEGLA 1021
Qy 1063 -----EAPRSLAPSPG-----AGSDVFDGDLGMAAGKGLQSLPT 1097
Db 1022 TSLGALSPLTGTLTRPGSGLSLSPSGYPMQSSLGACLDLAVLGGREQFSRPLSL 1081
Qy 1098 HDPSPQRYSEDPVPLPSETDGYV-----APL-----TC-----SQPE-----YVNPQ 1137
Db 1082 H-PIPRGR-----PASESEGHVGTGSEAELEQKVSVCRRSRSPRPRGDSAYHSQR 1133
Qy 1138 DVQPQSPRGRP-----LPAARPAAGATLERAKTILSP-GKNGV-----KDVF 1179
Db 1134 HSLLTPTVPLPSPGLEEBEDGNGYMPDTHLRGASSRREGTLSSVGLSVLGTBEDEED-- 1191
Qy 1180 AFGGAVENPEYLTPOGGAPOPHP 1204
Db 1192 -----EYEYNNRRKRGSP-PRPP 1209

RESULT 11
TVFVLV
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian leukosis virus
N;Contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase
C;Species: avian leukosis virus, ALV
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 11-Jun-1999
C;Accession: B00643; A00643
R;Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1985
A;Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A;Reference number: A00643; MUID:85228222; PMID:2988784
A;Accession: B00643
A;Molecule type: mRNA
A;Residues: 1-698 <NII>
A;Cross-references: GB:M10066; GB:M13881; NID:g211749; PIDN:AAA48763.1; PID:g211750
A;Note: In Genbank entry CHKERBBF, release 109.0, the source is designated as Gallus gal
C;Comment: This protein is synthesized as a gag-env-erbB protein.
C;Genetics:
A;Gene: gag-env-erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific p
F;1-6/Product: gag protein (fragment) #status predicted <GAG>
F;7-59/Product: env protein (fragment) #status predicted <ENV>
F;160-698/Product: protein-tyrosine kinase erbB #status predicted <ERB>
F;194-459/Domain: protein kinase homology <KIN>
F;202-210/Region: protein kinase ATP-binding motif
F;229/Active site: Lys #status predicted

Query Match 25.4%; Score 1723.5; DB 1; Length 698;
Best Local Similarity 51.5%; Pred. No. 3.6e-64;
Matches 368; Conservative 74; Mismatches 144; Indels 129; Gaps 17;
Qy 578 GPEADQCVACHYKDPFPCVARCPSGVKPDLSYMPIWPKFPEEGACQPCPNCNTHSCVDL 637

Db 60 GP--DHCMCKAHFIDGPHCVKACPAAGVLGENDTL-VWKYADANAVCOLCHPNCTRCKGCP 116
Qy 638 DDKGCPEAEQASPLTSIQYIKANSKF-----IGITELKRRQOKIRKYTMRRLLIQETELV 691
Db 117 GLEGCP---NGSKTPSIAAGVVGGLLCLVVVGIGLYLARRHIVRKRTLRLLQERELV 173
Qy 692 EPLTPSGAMINQOMRILKSTELRKVKVLGSGAGFTVYKGIWIIPDGENVKIPVAIKVLE 751
Db 174 EPLTPSGEAPNOAHLRIKTEFKFKVKVLGSGAGFTVYKGLWIPGEKVKIPVAIKLE 233
Qy 752 NTSPKANKEILDRAYVMAGVSPVSRLLGICLTSTVOLVQLTQMPYGCILLDHVRNRRGL 811
Db 234 ATPKANKKEILDRAYVMASVDNPHVCKLLGICLTSTVOLVQLTQMPYGCILLDHVRNRRGL 293
Qy 812 GSQDLLNWCQIAKMSYLEDVRLVHRDLAARNVLKSPNPKITDFGLARLLDIDETEH 871
Db 294 GSQVLLNWCQIAKMWYLBERRLVHRDLAARNVLKTPQHVKITDFGLAKLLGADEKEY 353
Qy 872 HADGGKVPKIKWMALESILRRRFTHQSDVMSYGVTVWELMTFCAGPYDGIIPAREIPDLLEK 931
Db 354 HADGGKVPKIKWMALESILHRIYTHQSDVMSYGVTVWELMTFCAGPYDGIIPAREIPDLLEK 413
Qy 932 GERLPOPICTIDVYIMVWKWMDSECRPRFRELVSFMRARDPQFVVIQ-NEDLGP 990
Db 414 GERLPOPICTIDVYIMVWKWMDSECRPRFRELVSFMRARDPQFVVIQ-NEDLGP 473
Qy 991 ASPLDSTFYRSLLEDVDAEYLVPOQGFPCDPAPGAGGMVHRHRSSTSRSGG 1050
Db 474 PSTDSKYFTLMEEDMEDIVDAEYLVPOQGF-----NPFST----- 513
Qy 1051 GDLTLGLEPSEEBAPRSL-----APSEGAGSVDFDGLGMAAGKGLQSLTPTDPSPLQR 1105
Db 514 -----SRTPLLSLSATSNNSATNCID-----RNGQGHVPVREDSFVQR 551
Qy 1106 YSEDPVPLPSET--DGYVAPLTCSPQPEYVQPDVVRPQPPSPREGPLPAARPAATLER 1163
Db 552 YSSDPTGNFLUESIDDGFL-----PAPEYVQ--LMPKKPS----- 585
Qy 1164 AKTLSPGKNGVWKDVF-----AFGGAVENPEYLTPOGGAAPQPHPPAFS 1208
Db 586 ---TAMVQNIYNNISITAIKLPMSRYQNSHSTAVDNEYL-----NTNQSPAK 634
Qy 1209 PAFDNLVYWDQ-----DPPE-----RGAPPSTFKGTPTAENPEYLGLDVP 1248
Db 635 TVFESSPYWTQSGNHQINLDNPDYQDQFLPNETKPNGLKLVKPAENPEYLRVAAP 689

RESULT 12
TVYUHV
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain H)
C;Species: avian erythroblastosis virus
C;Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 11-Jun-1999
C;Accession: A00644; A38022
R;Yamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.
Cell 35, 71-78, 1983
A;Title: The erbB gene of avian erythroblastosis virus is a member of the src gene family
A;Reference number: A00644; MUID:84026539; PMID:6313229
A;Accession: A00644
A;Molecule type: DNA
A;Residues: 1-604 <YAM>
A;Cross-references: GB:K01216; NID:g209676; PIDN:AAA42400.1; PID:g209678
R;Debuire, B.; Henry, C.; Benaisa, M.; Biserte, G.; Claverie, J.M.; Saule, S.; Martin, J.
Science 224, 1456-1459, 1984
A;Title: Sequencing the erbA gene of avian erythroblastosis virus reveals a new type of
A;Reference number: A38022; MUID:84223957; PMID:6328658
A;Accession: A38022
A;Molecule type: DNA
A;Residues: 1-28,'W',30-139,'F',141-145,'V',147-152 <DEB>
A;Cross-references: GB:K02006
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific p

F:130-395/Domain: protein kinase homology <KIN>
F:138-146/Region: protein kinase ATP-binding motif
F:165/Active site: Lys #status predicted

Query Match 24.4%; Score 1660; DB 1; Length 604;

	Qy	587	CAHYKDPDFCVARCPGKVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPABEQ	646
	Db	3	CAHPIDGPHCKVACAPAGVLGENDTL-VRKYADANAVCQLCHPNCITRGCKGPLEGCP---	58
	Qy	647	RASPLTSTIOYIKANSKF-----IGITELKRROOKIRKYTWRRLLQETVELPELTPSGAM	700
	Db	59	NGSKTPSTAAGVVGGLCLVVVGIGIGLYLRHRHVIRKTRRLRLLQERLEVEPLTPSGEA	118
	Qy	701	PNQAOMRIKETELRKVKVLGSGAFGTGVYKGIWIPDGENVKIPIVAIKVLRENTSPKANKE	760
	Db	119	PNQAHLRLIKETEKKFKVKVLGSGAFGTGYKGLWIPEGEKVKIPIVAIKELREATSPKANKE	178
	Qy	761	ILDEAYVMAGVGSPPVSRLLGICLTSTVOLVTOLMPYGCILLDHVRENRGLSGDLLNWC	820
	Db	179	ILDEAYVNASVDNHVCVRLGLCLTSTVOLITQOLMPYGCLLDYIREHKDNJISQYLNNWC	238
	Qy	821	MQIAKMSYLEDRVLVHRDLAARNLVKS PNHVKITDFGLARLLDIDETEHADGGKVP I	880
	Db	239	VQIAKGMNYLEERRLVHRDLAARNLVKTPOHVKITDFGLAKLGADKEYHAEGKVPI	298
	Qy	881	KWMALESILRRRFHQSDVNSYGTVWELMTFGAKPYDGI PAREYPDLLXKERLPOPPI	940
	Db	299	KWMALESILHRIYTHQS DVNSYGTVWELMTFGSKPYDGI PASEISSVLEKGERLPQPP I	358
	Qy	941	CTIDVIMIMVKCMIDSECRPRFRELVSFEFSRMARDPQRFVVIQ--NEDLGPASPJDSTFY	999
	Db	359	CTIDVIMIMVKCMIDADS RPKFERLIAEFSKMARDPPRYLVIOGDERMHLPSPTDSKFY	418
	Qy	1000	RSILLEDDMGDLVDAEEYLVFOOQFFCDPPAPGACGMVHHRRSSSTS RGGGDLLGLEP	105
	Db	419	RTLMEEDMEDIVDAEYLVPHQGPF-----NSPST-----	449
	Qy	1060	SEEBAPSPL-----APSEGAGSDVFDDGLGMAAKQLSLPTHDPSP LQRYSDEPTVPL	1111
	Db	450	-----SRTPLLSSLATSNNSATNCID-----RNQGHVPREDSFQVRYS S DPTGNF	496
	Qy	1115	PSET--DGYYAPLTCSPQPEYVNOQDPVRQPQPPSPREGPLPAARPA GATLERAKT LUSPGKN	1177
	Db	497	LEESIDDGFL-----PAPEYVNO--LMPKKPSTAM-----	524
	Qy	1173	GVRKDVRFAF-----GGAYENPEYILTPOGGAAPQHPPHPAFSPADF N	1211
	Db	525	--VQNQIYNFISLTAISKLPMDSR YQNSHSTA VDNFEYL-----NTNQSP LAKTVFES	575
	Qy	1214	LYYWDQDPPERGAPPSTTKGTPTAE NPEY 1242	
	Db	576	SPYIIQSGNHQ-----INLNPDY 594	

Nature 314, 178-180, 1985

A;Title: A *Drosophila* genomic sequence with homology to
 Nature 314, 178-180, 1985
 A;Reference number: A38021; MUID:85137938; PMID:2983232

A;Accession: A38021

A:Molecule type: DNA

A, molecule type. DNA
A:Residues: 'A', 832-866, 'V', 868-943, 'OTPSLVK' <WAD>

A:Accession: A1952800, V1900913, Z1900914
A:Cross-references: EMBL:X02993; NID:g7922; PIDN:CAA26157.1; PID:g929565
C:Comment: This sequence is tentative because the introns have not been identified.

A:Gene: FlvBase: Eafr

A:Gene: FlyBase:FBgt
A:Cross-references: FlyBase:FBan0003731

A:Man position: 2 57E
A:Cross-references: F

A;Map position: 2 3/F
C:Superfamily: epidermal growth factor receptor: protein kinase homology

[illegible]

```
C;keyWords: A1F; autoPhosphorylation; duplication; glycosylation;
E:1-732/Domain: extracellular #status predicted <EXT>
```

F;I-/32/Domain: extracellular #status predicted <EX>
E;733-764/Domain: transmembrane #status predicted <TM>

```

E:/333-64/domain: transmembrane #status
E:365-1330/domain: intracellular #status

```

F;765-1330/Domain: intracellular #status predicted <INT>
F;808-1072/Domain: protein kinase homolog <KTN>

F;808-1072/Domain: protein kinase homology <KIN>
F;816-824/Region: protein kinase ATP-binding motif

Query Match 23.8%; Score 1617.5; DB 1; Length 1330;

Db 655 QRECFQRPEC---NG---CTGPGADCKSRNFKLFDFANETGYVNSTMFCNCKSKPLE 708
Qy 604 VK-PDLSYMPYKFPDEBACQPCPINCTHSCVDLDDKCPABORASPLTSIQYIKANSK 662
Db 709 MRHVNQYATLGPY-----CAASP-----PRSSKITA--NLDVNM 742
Qy 663 FI-----GITEKRRQOKIRKYT--MRRLLQETELVEPLTPSGAMPNQAQM 706
Db 743 FIITGAVLVPTICILCVVYTCRQKAKKQAKVETKMTWALSREDSEPLRPSNIGANLCKL 802
Qy 707 RILKETELRVKVLGSGAGFYVGIWIPDGENVKIPVAIKVLENTSPKANKILDEAY 766
Db 803 RIVDAELRKGVLGSGAGFYVGIWIPDGENVKIPVAIKVLENTSPKANKILDEAY 862
Qy 767 VMAGVGSPPYRLLGICLTSTVOLTPMPYVGLLHVRENRLGSDLLNWCWQIAKG 826
Db 863 IMASEEHVNLKLLAVCMSSOMLITQMPYLGCLLDVYRNRDKIGSKALLNWTQAKG 922
Qy 827 MSYLEDVRLVHRDLAARNVLK---SPNHVKITDFGLARLLDIDETEYHADGGKVPKIM 883
Db 923 MSYLEERLRLVHRDLAARNVLVRLLAGEDH---DFGLAKLLSSDNEYKAAGGKMPKWL 978
Qy 884 ALESILRRRPTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPICTI 943
Db 979 ALECIIRRVFTSKSDVWAFGTIWEILLTFQRPHEHIPAKDIPDLIEVGLKLEQPEICSL 1038
Qy 944 DVYIMVWKMWIDSECRPRFRELSEFSRMDARDPQRFVVIQNEDLG--PASPLDSTFYS 1001
Db 1039 DIYCTLLSCHWLDAAKPTFKQITVFAEPFARDPGRYLAIGDKFTRLPA-----YTS 1091
Qy 1002 LLED--DMGDLVDAEYLVPOQGFCCPDPAAGAGGMVHRRSSSTRSGGDLTLGLE 1058
Db 1092 QDEKDLIRKLAPTGDSEAIKPDYILQPKAALGPS-----HRTDCT----- 1133
Qy 1059 PSEEEAP-----RSPLAPSEGAGSVFDG---DLGMAAGKLGSLPHTHDPSPLORYSED 1109
Db 1134 ---DEMPKLNRYCKDPNKNKNSSTGDDERDSSAREVGVGNLR----- 1171
Qy 1110 PTVPLPSETDGYVAPLTCSPQPEYVNOPDVRPQPPSPREGPLPAARPAAGATLERAKTLSP 1169
Db 1172 --LDLPVDEDDYLMPT-TCQCPGNNNNNN-----NPNQNNMAAVGVAAGYM----- 1214
Qy 1170 GKNGVVKDVFAGGAVENPEYL-----TPQGAAPQPH-----PPPAF 1207
Db 1215 -----DLIGVPVSDNPEYLLNAQTLGVGESPIQTIGIPVMGGPGTMEVKVPMG 1267
Qy 1208 SP-AFDNLYWD 1218
Db 1268 EPTSSDHEYND 1279

RESULT 14

S35745
C:Species: kinase-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus
C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific P
C:Date: 03-Mar-1994 #sequence_revision 26-May-1995 #text_change 28-Feb-1997
C:Accession: S35745
R:Vennstrom, B.
Submitted to the EMBL Data Library, March 1993
A:Reference number: S35743
A:Accession: S35745
A:Molecule type: DNA
A:Residues: 1-544 <VEN>
A:Cross-references: EMBL:X12707
C:Genetics:
C:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific P
F:135-400/Domain: protein kinase homology <KIN>
F:143-151/Region: protein kinase ATP-binding motif
F:170/Active site: Lys #status predicted

Query Match 23.6%; Score 1604; DB 2; Length 544;
Best Local Similarity 54.1%; Pred. No. 2.3e-59;
Matches 339; Conservative 64; Mismatches 128; Indels 96; Gaps 14;
Qy 578 GPEADQCVACAHYKDPFPCVACPSGVKPDLSYMPYKFPDEBACQPCPINCTHSCVDL 637
Db 1 GP--DHCMKCAHFDGPHCVKACPAAGVLGENDTL-VWKYADANAVCOLCHPNCNTRGCKGP 57
Qy 638 DDGCGPAEQRAJASPLTSIQYIKANSK-----IGITELKRRQOKIRKYTMRRLLOETELV 691
Db 58 GLEGCP---NGSKTPSIAAGVVGGLLCLLVVVGILGLYLRRLHVRKTLRLLOETELV 114
Qy 692 EPIPTSGAMPNQAQMILKETELRVKVLGSGAGFYVGIWIPDGENVKIPVAIKVLE 751
Db 115 EPLTPSGEAPNQAHLRIKETEFKVKVLFGAGFYVGIWIPDGENVKIPVAIKVLE 174
Qy 752 NTSPEKANKETILDAYVMAGVGSPPVSRLLGICLTSTVOLTPMPYVGLLHVRENRLG 811
Db 175 ATSPKANKETILDAYVMAGVGSPPVSRLLGICLTSTVOLTPMPYVGLLHVRENRLG 234
Qy 812 GSODLLNWCWQIAKMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEY 871
Db 235 GSQYLLNWCWQIAKMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEY 294
Qy 872 HADGGKVPKIMWALESILRRRPTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEK 931
Db 295 HAEGKVPKIMWALESILRRRPTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEK 354
Qy 932 GERLPQPICTIDYIMVWKMWIDSECRPRFRELSEFSRMDARDPQRFVVIQNEDLG 990
Db 355 GERLPQPICTIDYIMVWKMWIDSECRPRFRELSEFSRMDARDPQRFVVIQNEDLG 414
Qy 991 ASPLDSTFYRSLLEDMDGDLVDAEYLVPOQGFCCPDPAAGAGGMVHRRSSSTRSGG 1050
Db 415 PSTDSKFTYLTNEEDMEDIVDAEYLVPOQGF-----NSPST----- 454
Qy 1051 GDLTLGLEPSEEBAPRSP-----APSEGAGSVFDGDLGMAAGKLGSLPHTHDPSPLO 1105
Db 455 -----SRTPLLSLSATSNNSATNCIDRNGG-----H----- 481
Qy 1106 YSEDPTVPLPSETDGYVAPLTCSPQPEYVNOPDVRPQPPSPREGPLPAARPAAGAT-LERA 1164
Db 482 -----PVREDGFL-----PAPEYVNOQ--LMPKPKSTAMVQNIYVLSLTAKSL 524
Qy 1165 KTLSPGKNGVVDVFAFGGAVENPEYL 1191
Db 525 PIDSRVQN-----SHSTAVDNPEYL 544

RESULT 15

S00727
C:Species: kinase-related transforming protein (erbB) (EC 2.7.1.1-) - avian erythroblastosis virus
C:Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 28-Feb-1997
C:Accession: S00727
R:Scotting, P.; Vennstrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.
Oncogene Res. 1, 265-278, 1987
A:Title: Common site of mutation in the erbB gene of avian erythroblastosis virus mutant
A:Reference number: S00727; MUID:88217326; PMID:2897102
A:Accession: S00727
A:Molecule type: DNA
A:Residues: 1-545 <SCO>
A:Cross-references: EMBL:X06943
C:Genetics:
C:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; phosphotransferase
F:135-400/Domain: protein kinase homology <KIN>
F:143-151/Region: protein kinase ATP-binding motif

Query Match 23.5%; Score 1597; DB 2; Length 545;
Best Local Similarity 54.1%; Pred. No. 4.6e-59;
Matches 339; Conservative 63; Mismatches 129; Indels 96; Gaps 14;

```
QY 578 GPEADQVACAHYKDPPECPVARCPGVKPDLSYMPIMKFPDEGACOPCPINCTHSCVDL 637
Db 1 GP--DHCWKCAHFDGPHCVACFAGVUGENDTL-VMKIADANAVCQLCHENCCTRGCKGP 57
QY 638 DDKCPAEQASPLTSIOYIKANSKF-----IGITELKRQKQKIRKYTMRRLLQETELV 691
Db 58 GLEGCP---NGSKTPSIAAGVVGGLCLVAVVGLGIGLYLRRRHIVRKRTLRLRLQERELV 114
QY 692 EPLTPSGAMPNOAQMRILKETELKRVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRE 751
Db 115 EPLTPSGEAPNOAHLRIKETEFKVKVVLGFGAFGTVYKGLWIPEGEKVTIPVAIKELRE 174
QY 752 NTSFKANKEIIDEAVVWAGVSPVSRLLGICLTSTVOLVTLQMPYGCGLDHDVRENRL 811
Db 175 ATSPKANKEIIDEAVVWASVUNPHVCRLLGICLTSTVOLITQLMPYGCCLDIYREHKONI 234
QY 812 GSQDLLNMCQIAKGMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEY 871
Db 235 GSQYLLNWCVOIAGMNVLEERHLVHRDLAARNVLKTPQDVKITDFGLAKQLGADEKEY 294
QY 872 HADGKVPKIMWALESLIRRRFTQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEK 931
Db 295 HAEGGKVPKIMWALESLIHRITYHQSDVWSYGVTVWELMTFGSKPYDGIPIASEISSVLEK 354
QY 932 GERLPQPPICITIDVYIMVWCMWIDSECRPRFRELVSFEFSRMARDPORFVVIQ-NEDLGP 990
Db 355 GERLPQPPICITIDVYIMVWCMWDADSRPKFRELIAEFKWARDPPRYLVIOGDERMHL 414
QY 991 ASPLDSTFYRSLLEDDDDGLVDAEYLVPOQGFPCPDPAFGAGMVHRRHRSSTRSGG 1050
Db 415 PSPTDSKIFYRTLMBEEDMEDIVDAEYLVPHQGF-----NSPST--- 454
QY 1051 GDLTGLPESEEBAPRSP-----APSEGAGSDVFDGLGMAAKGLQSLPTHDPSPLO 1105
Db 455 -----SRTPLLSLSATSNNSATNCIDRNG-----H----- 481
QY 1106 YSEDPTVPLPSETDGYVAPLTCSPQPEYVNOVDYRPOPPSPREGPLPAARPAGAT-LERA 1164
Db 482 -----PVREDFL-----FAPEYVNO--LMPKKPSTAMVQNIYNYISLTAISKL 524
QY 1165 KTLSPGKNGVVKDVFAGGAVENPEYL 1191
Db 525 PMDSRYQN-----SHSTAVDNPEYL 544
```

Search completed: July 22, 2003, 09:27:20
Job time : 28.179 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 09:10:23 ; Search time 22.1435 Seconds
(without alignments)
5413.772 Million cell updates/sec

Title: SEQ4-653-675-12
Perfect score: 6776
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYGLDVPV 1247
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6602	97.4	1255	1 A24571	protein-tyrosine k
2	5809	85.7	1260	1 TVRTNU	protein-tyrosine k
3	5799.5	85.6	1254	2 I48161	p-185 precursor -
4	3066	45.2	1210	1 GQHUE	epidermal growth f
5	3032	44.7	1210	2 A53183	epidermal growth f
6	3011.5	44.4	1223	1 TVCHLV	epidermal growth f
7	2915.5	43.0	1308	2 A47253	epidermal growth f
8	2596	38.3	1166	1 S06142	protein-tyrosine k
9	2381.5	35.1	1342	2 A36223	kinase-related tra
10	2293.5	33.8	1339	2 JC4387	epidermal growth f
11	1683.5	24.8	698	1 TVFVLV	protein-tyrosine k
12	1620	23.9	604	1 TVYUHV	protein-tyrosine k
13	1612.5	23.8	1330	1 GQFFE	epidermal growth f
14	1564	23.1	544	2 S35745	protein-tyrosine k
15	1557	23.0	545	2 S00727	kinase-related tra
16	1540	22.7	540	2 B44776	protein-tyrosine k
17	1538	22.7	540	1 TVFVFB	protein-tyrosine k
18	1509	22.3	644	2 A36325	epidermal growth f
19	1287	19.0	1323	2 E88257	protein let-23 (im
20	1287	19.0	1374	2 S70712	protein-tyrosine k
21	1197	17.7	1369	2 S70713	protein-tyrosine k
22	1144	16.9	1717	1 A45558	epidermal growth f
23	1126	16.6	527	2 A42032	epidermal growth f
24	975.5	14.4	843	2 A27131	epidermal growth f
25	806.5	11.9	346	2 S13807	protein-tyrosine k
26	754.5	11.1	311	2 S13808	protein-tyrosine k
27	735	10.8	1363	2 T43220	insulin-like growth
28	704	10.4	1383	2 A36080	insulin receptor p
29	703.5	10.4	1372	2 A34157	insulin receptor p

ALIGNMENTS

RESULT 1

A24571

Protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human

N/Alternate names: C-erb-B-2 protein precursor; kinase-related transforming protein erbB
C/Species: Homo sapiens (man)
C/Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999
C/Accession: A24571; A25491; A44188; B44188; I59509; I57522
R/Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T.;
Nature 319, 230-234, 1986
A/Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth f
A/Reference number: A24571; MUID:86118663; PMID:3003577
A/Accession: A24571
A/Molecule type: mRNA
A/Residues: 1-1255 <YAM>
R/Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.
Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985
A/Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epiderm
A/Reference number: A25491; MUID:86016729; PMID:2995967
A/Accession: A25491
A/Molecule type: DNA
A/Residues: 737-1031 <SEM>
A/Cross-references: GB:M11767; NID:g182163; PIDN:AAA35808.1; PID:g553282
R/Coussens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg, P.
Science 230, 1132-1139, 1985
A/Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chromos
A/Reference number: A44188; MUID:86070181; PMID:2999974
A/Accession: A44188
A/Molecule type: DNA
A/Residues: 740-910 <COU1>
A/Cross-references: GB:M12036; NID:g183988; PIDN:AAA35978.1; PID:g183989
A/Accession: B44188
A/Molecule type: mRNA
A/Residues: 1-517, 'RALL', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>
A/Cross-references: GB:M11730; NID:g183986
R/King, C.R.; Kraus, M.H.; Aaronson, S.A.
Science 229, 974-976, 1985
A/Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.
A/Reference number: I59509; MUID:85272597; PMID:2992089
A/Accession: I59509
A/Status: translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 832-909 <REX>
A/Cross-references: GB:I29395; NID:g459807; PIDN:AAA35809.1; PID:g459808
R/Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.
Mol. Cell. Biol. 7, 2597-2601, 1987
A/Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptional
A/Reference number: I57622; MUID:87286898; PMID:3039351
A/Accession: I57622
A/Status: translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-191 <TAL>

insulin receptor p
insulin receptor-r
insulin-like growth
protein-tyrosine k
insulin receptor-r
insulin-like growth
insulin-like growth
insulin-like growth
insulin receptor -
insulin receptor -
insulin receptor -
protein-tyrosine k
protein-tyrosine k
protein-tyrosine k
protein-tyrosine k
mouse developmenta

A;Cross-references: GB:M16792; NID:g183983; PIDN:AAA58637.1; PID:g553332
C;Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
C;Genetics: GDB:ERBB2; NGL; NEU; HER-2
A;Gene: GDB:ERBB2; NGL; NEU; HER-2
A;Cross-references: GDB:120613; OMIM:164870
A;Map position: 17q21.1-17q21.1
A;Introns: 25/1; 75/3; 147/1; 883/3
A;Note: the list of introns is incomplete
C;Function:
A;Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
inase
F;1-21/Domain: signal sequence #status predicted <SIG>
F;122-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>
F;122-653/Domain: extracellular #status predicted <EXT>
F;70-304/Domain: EGF receptor extracellular domain repeat <EE1>
F;395-605/Domain: EGF receptor extracellular domain repeat <EE2>
F;654-675/Domain: transmembrane #status predicted <TM>
F;676-1255/Domain: intracellular #status predicted <INT>
F;718-983/Domain: protein kinase homology <KIN>
F;726-734/Region: protein kinase ATP-binding motif
F;68.124,187,530,571,629/Binding site: carboxylate (Asn) (covalent) #status predicted
F;686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F;753/Active site: Lys #status predicted
F;1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 97.4%; Score 6602; DB 1; Length 1255;
Best Local Similarity 97.2%; Pred. No. 5.3e-265;
Matches 1220; Conservative 9; Mismatches 18; Indels 8; Gaps 3;

Qy 1 MELAALCWGULLALLPPGAASVCTCTDKMLRLPASPEHLMRLHYOGCQVQGNL 60
Db 1 MELAALCWGULLALLPPGAASVCTCTDKMLRLPASPEHLMRLHYOGCQVQGNL 60
Qy 61 ELTYLPTNASLSFLQIDQEVQYVLIHANNQVQPLQLRLVRGTLQFEDNYALAVLNG 120
Db 61 ELTYLPTNASLSFLQIDQEVQYVLIHANNQVQPLQLRLVRGTLQFEDNYALAVLNG 120
Qy 121 DPLNNTTPTVGASPGGLRELQRLSLEILKGVLIQVLPOLCYQDTILKWDIFHNQOLA 180
Db 121 DPLNNTTPTVGASPGGLRELQRLSLEILKGVLIQVLPOLCYQDTILKWDIFHNQOLA 180
Qy 181 LTLIDTNSRACHPCSPCKSRGWSESDCOSITRTVCAGGARCKGPLEPTCCHEQC 240
Db 181 LTLIDTNSRACHPCSPCKSRGWSESDCOSITRTVCAGGARCKGPLEPTCCHEQC 240
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESNPNEGRTYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESNPNEGRTYTFGASCVTACP 300
Qy 301 NYNLSTDVGSCTLVCLNHNQVTSADGTQRCCKSPCARVCYGLMGQYIKANSKFIQIT 360
Db 301 NYNLSTDVGSCTLVCLNHNQVTSADGTQRCCKSPCARVCYGLMGQYIKANSKFIQIT 360
Qy 361 ELEFAGCKKIFGSLAFLESFEDGDPASNTAPLOEQLOVFTLEITCYLYISAWPDSLP 420
Db 361 IOEFAGCKKIFGSLAFLESFEDGDPASNTAPLOEQLOVFTLEITCYLYISAWPDSLP 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLISWGLRLSRLRELGSGLALHNNHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLISWGLRLSRLRELGSGLALHNNHLCFVHTV 480
Qy 481 PWDQFRNPHOALLTANRPEDECVGEGGLACHQLCARGHCWGPQTQCVCNSQFIRGQEC 540
Db 481 PWDQFRNPHOALLTANRPEDECVGEGGLACHQLCARGHCWGPQTQCVCNSQFIRGQEC 540
Qy 541 VEECVLQGLPREYNARHCLPCHPECOPOGNSVTCFGEADOCVACAHYKDPPECVARC 600
Db 541 VEECVLQGLPREYNARHCLPCHPECOPOGNSVTCFGEADOCVACAHYKDPPECVARC 600
Qy 601 PSGVKPDLSYMPKIFPDEEGACQPCPINCTHSCVDLDDKGCQPAEQRASPLTISIYAVVG 660
Db 601 PSGVKPDLSYMPKIFPDEEGACQPCPINCTHSCVDLDDKGCQPAEQRASPLTISIYAVVG 660

Db 601 PSGVKPDLSYMPKIFPDEEGACQPCPINCTHSCVDLDDKGCQPAEQRASPLTISIYAVVG 660
Qy 661 ILLVVLGVVFGILJ-----QYIKANS--KFIGITEL--PLTPSGAMPNQAMRILKETEL 712
Db 661 ILLVVLGVVFGILJKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
Qy 713 RKVKVLGSGAGFTGVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKILDEAYVNAVVGSP 772
Db 721 RKVKVLGSGAGFTGVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKILDEAYVNAVVGSP 780
Qy 773 YVSRLLGICLTSTVQLVTLMPYGLCLLDHVRNRLGSLQDLNLCMQOIAKMSYLEDV 832
Db 781 YVSRLLGICLTSTVQLVTLMPYGLCLLDHVRNRLGSLQDLNLCMQOIAKMSYLEDV 840
Qy 833 LVHRDLAARNVLKSPNHNKIDTGLARLLDIDETEHADGGKVPKIMMALESILRRRFT 892
Db 841 LVHRDLAARNVLKSPNHNKIDTGLARLLDIDETEHADGGKVPKIMMALESILRRRFT 900
Qy 893 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVVKCWM 952
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVVKCWM 960
Qy 953 IDSECRPRFRELVSFSEFARMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDDMGDLVDA 1012
Db 961 IDSECRPRFRELVSFSEFARMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDDMGDLVDA 1020
Qy 1013 EYILVPOQGFCDPPAPGAGMWHHRSSSTRSGGDLTLGLEPSEEEAPRSPPLAPSEG 1072
Db 1021 EYILVPOQGFCDPPAPGAGMWHHRSSSTRSGGDLTLGLEPSEEEAPRSPPLAPSEG 1080
Qy 1073 AGSDVDFGDLGMAAKQLSLPHTDPSFLORYSDPTVPLPSETDGVVAPLTCSPQPEYV 1132
Db 1081 AGSDVDFGDLGMAAKQLSLPHTDPSFLORYSDPTVPLPSETDGVVAPLTCSPQPEYV 1140
Qy 1133 NQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGXGVVYKVFAGGAVENPEYLTPO 1192
Db 1141 NQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGXGVVYKVFAGGAVENPEYLTPO 1200
Qy 1193 GGAAPQHPHPPAPAFDNLVYWDQDPPERGAPSTFKGTPTAENPEYGLDVV 1247
Db 1201 GGAAPQHPHPPAPAFDNLVYWDQDPPERGAPSTFKGTPTAENPEYGLDVV 1255

RESULT 2
TVRNU
protein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 11-Jun-1999
C;Accession: A24562; A61204
R;Bargmann, C.I.; Hung, M.C.; Weinberg, R.A.
Nature 319, 226-230, 1986
A;Title: The neu oncogene encodes an epidermal growth factor receptor-related protein.
A;Reference number: A24562; MUID:86118662; PMID:3945311
A;Accession: A24562
A;Molecule type: mRNA
A;Residues: 1-1260 <BAR>
A;Cross-references: EMBL:X03362; NID:g56745; PIDN:CAA27059.1; PID:g56746
R;Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smath, R.A.; Cohen, Carcinogenesis 12, 1975-1978, 1991
A;Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals no m 2-thiazolyl]formamide or N-methyl-N-nitrosourea.
A;Reference number: A61204; MUID:92035293; PMID:1682063
A;Accession: A61204
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 637-663,'V',665-702 <MAS>
A;Note: authors translated the codon GCA for residue 25 as Val
C;Genetics:
A;Gene: neu
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>

F:658-680/Domain: transmembrane #status predicted <TN>
F:723-988/Domain: protein kinase homology <KIN>
F:731-739/Region: protein kinase ATP-binding motif
F:71,191,263,535,576,634/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:691/Binding site: phosphate (Thr) (covalent) #status predicted
F:758/Active site: Lys #status predicted
F:802,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

Qy	1	MELAAALCRWGLLLALLPPGAAS	TQVCTGDMKRLR	PAS	PETHLDMRLHLYQGQGVQGNL	60															
Db	4	MELAAACRWGFLALLPPGIAGTQVCTGDMKRLR	PAS	PETHLDMRLHLYQGQGVQGNL	63																
Qy	61	ELTYLPTNASLSFLQDIQSVQGVLI	AHNOVQVPLQR	LIRVGTOL	FEDNYALAVLDNG	120															
Db	64	ELTYVPAVASLSFLQDIQSVQGM	LIAHNOVKRVLQR	LIRVGTOL	FEDKYALAVLDNR	123															
Qy	121	DPLNNTTPTV	TGASPGGLRE	LQRLS	TEILKGGVLITQRP	QCYQDTILWKDIFHKNNQL	179														
Db	124	DPQDNVAASTPGRT	PEGLRELQRLS	TEILKGGVLIRGN	PQCYQDWLWKDIFVRKNNQL	183															
Qy	180	ALTLDITNRSRACHPCSP	CMCKSGSR	CWGESSEDCQS	LTRVCAAGCARCKG	PLPTDCHEQ	239														
Db	184	APVDITNRSRACPP	CAPACKDNHCWGESPE	DCQILTGTICTSG	CARCKGRLPTDCHEQ	243															
Qy	240	CAAGCTGPKHSDCLAC	LHFNHSGI	CELHCPALVYNT	NTDTFESMNP	EGRYTFGASCVTAC	299														
Db	244	CAAGCTGPKHSDCLAC	LHFNHSGI	CELHCPALVYNT	NTDTFESMNP	EGRYTFGASCVTTC	303														
Qy	300	PYNYLSTDVGSCT	LVCPHLNQVTAEDGTQ	RCCKSPCARVCYGL	GMQVIKANSK	FIGI	359														
Db	304	PYNYLSTEVGSC	TLVCPNNQVTAEDGTQ	RCCKSPCARVCYGL	GMELHRCARITSD	363															
Qy	360	TELEFAGCKKIFGS	LAFLPESFDGDPAS	NTAFLQPEQLQVF	LEITGYLISAMP	PDSL	419														
Db	364	NVQEFDGCKKIFGS	LAFLPESFDGDPSS	GIALPLRPQQLQVF	LEITGYLISAMP	PDSL	423														
Qy	420	PDLVSFQNLQVIR	GIHLNGAVSLTLOG	LIGLSWLGRLS	RELGSGLALIH	NTHLCFVHT	479														
Db	424	RDLVSFQNLRIIR	GIHLHDGAVSLTQ	LIGLHLSGLRS	RELGSGLALIH	RNAHLCFVHT	483														
Qy	480	VPDQLFRNP	HOALLHTANR	PRDE-CV	GBGLACHOLCARG	HCWGPPTOCVNC	SQFLRGQ	538													
Db	484	VPDQLFRNP	HOALLHSGNR	PREDIC	LVSSGLVCSNLCAH	GCWGPPTOCVNC	SHFRLRGQ	543													
Qy	539	ECVEECRV	LOGLPRYVNA	RHCLPCHPE	CPQNGSVTCF	GP	EADQC	VACAHYKDPFCVA	598												
Db	544	ECVEECRV	WKGLPREYVSD	KRLCPHCE	CPQNSSETCF	GS	EADCAACAHYK	SDSSCVA	603												
Qy	599	RCPSGVK	PDLVSMP	TPKPEDE	GCACQPCD	INCHTSCVD	LDKCCP	AEORASPLTS	IVSAV	658											
Db	604	RCPSGVK	PDLVSMP	TPKPYDE	EGICQPCD	INCHTSCVD	LDKCCP	AEORASPT	FTFIATV	663											
Qy	659	VGILLVVL	GVVFGIL	---QV	IKANS--K	FIGITEL--	PLTPSGAMP	NOAQMR	ILKET	710											
Db	664	EGVLLFL	VLVVVG	LILKRR	QKIRKYTWRR	LLQETEL	VEPLTPSGAMP	NOAQMR	ILKET	723											
Qy	711	ELRKVK	VLGSGAF	GTVYK	GIWTPD	GENVKIPVA	IKVLR	ENTSP	KANKEIL	DEAYVMAGVG	770										
Db	724	ELRKVK	VLGSGAF	GTVYK	GIWTPD	GENVKIPVA	IKVLR	ENTSP	KANKEIL	DEAYVMAGVG	783										
Qy	771	SPYVSR	LLGICLT	STVQL	VTQMPY	GCLLD	HVRENR	GRIGSQ	DLNWC	QIAKMS	YLED	830									
Db	784	SPYVSR	LLGICLT	STVQL	VTQMPY	GCLLD	HVREH	GRIGSQ	DLNWC	QIAKMS	YLED	843									
Qy	831	VRLVHR	DLAARN	VLKSP	NHVKITD	FGLAR	LDDIDET	EYHAD	GKVP	IKWMA	LESILRR	890									
Db	844	VRLVHR	DLAARN	VLKSP	NHVKITD	FGLAR	LDDIDET	EYHAD	GKVP	IKWMA	LESILRR	903									
Qy	891	FTHOSD	VMSYGV	TVMEL	MTFG	AKPYD	GI	IPARE	IPD	LLK	GER	LP	Q	PI	CT	DI	VY	MT	VM	KVC	950

A:Molecule type: mRNA
A:Residues: 969-971, 'K', 973-1115, 'D' <BS>
A:Cross-references: EMBL:Z12608
R:Heisermann, G.J.; Gill, G.N.
J. Biol. Chem. 263, 13152-13158, 1988
A:Title: Epidermal growth factor receptor threonine and serine residues phosphorylated
A:Reference number: A28941; MUID:88330814; PMID:3130233
A:Accession: A28941
A:Molecule type: protein
A:Residues: 689-694, 'X', 696-704, 'L', 706-707, 989-992, 'XX', 995-996, 'X', 998-1000; 1002-1009,
R:Hibbs, M.L.; Dunn, A.R.; Alexander, W.S.
submitted to the EMBL Data Library, April 1994
A:Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor
A:Reference number: S45325
A:Accession: S45325
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-971, 'K', 973-1210 <VER>
A:Cross-references: EMBL:X78987; NID:G488030; PIDN:CAA55587.1; PID:G488831
R:Paria, B.C.; Das, S.K.; Andrews, G.K.; Day, S.K.
Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993
A:Title: Expression of the epidermal growth factor receptor gene is regulated in mouse
A:Reference number: 149643; MUID:93126380; PMID:7678348
A:Accession: 149643
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 12-20, 22-132 <RES>
A:Cross-references: GB:L06864; NID:G193001; PIDN:AAA53029.1; PID:G567201
C:Genetics:
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphoprotein
F:1-24/Domain: signal sequence #status predicted <SIG>
F:648-670/Domain: transmembrane #status predicted <TM>
F:712-977/Domain: protein kinase homology <KIN>
F:720-728/Region: protein kinase ATP-binding motif
F:680, 695/Binding site: phosphate (Thr) (covalent) #status experimental
F:697, 1070, 1071/Binding site: phosphate (Ser) (covalent) #status experimental
F:993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental
F:1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental
F:1197/Binding site: phosphate (Tyr) (covalent) #status experimental

Query Match 44.7%; Score 3032; DB 2; Length 1210;
Best Local Similarity 48.8%; Pred. No. 6, 3e-118;
Matches 621; Conservative 170; Mismatches 363; Indels 118; Gaps 26;
QY 11 LLLALLPPGAA--STQVCTGTDMLRLPASPTHLDMLRLHYQQGVQGVQGNLELYLPTN 68
DB 14 LTLALCAAGGAALEKKVCGQTSNRLTQLGTFFDHLFLSLQRMVNNCEVVLGNLEITYQVN 73
QY 69 ASLSFLQIOEQVGVVLAHNOVROVPLORLIVRGTOIFEDNYALAVLDNGDPLNNTTP 128
DB 74 YDLSEFKTIOEVAGVYVLAALNTVERIPLENQITREGNALYENTYALATLSN----- 124
QY 129 VTGASPGGLRELQRLSLTEILKGGVLIQKNPOLCYQDTILWKDI----FHKNNOLATLI 184
DB 125 -YGNRTGLRELPMNLQELLIGATVFNPNILCNWDIQRDIQVNVFMSNWSMDL---- 180
QY 185 DTRNSRACHPCSPMKGSRGSESSDCQSLTRTVACGCA-RCKGPLPTCCCHQCAAG 243
DB 181 -QSPSPSCPKCDPSPNGSCWGGGECQKLTIIQAQCSHRCGRSPSCCHNQCAAG 239
QY 244 CTGPKHSCLACLFHNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPNY 303
DB 240 CTGPRSDCLVCQKQFODBATCKDTCPLMLYNPTTYQMDVNPPEGKYSFGATCVKCKPRN 299
QY 304 LSTDVGSCTLVCLPHNQEVATBDGTFQCEKSKSPCARVCYGLGMQYIKANSKFIGITBLE 363
DB 300 VYDHGSCVRACGPDYEV-EDGIRKCKKCDGPKVCNGIGICEFK-DTILSNATNIK 357
QY 364 -FAGCKITGSLAFIPESFGDGPASNTAPLOEQLOVFETLEITGYLYISAWPDSLPDL 422
DB 358 *HPKYCTAISGDLHILFVAFKGSFTRTPPLDPRELEILKTVKEITGFLLIQAWPDNWTDL 417

RESULT 6

TVCHLV

Epidermal growth factor receptor precursor - chicken

N:Contains: protein-tyrosine kinase (EC 2.7.1.12) erbB

C:Species: Gallus gallus (chicken)

C:Date: 28-Feb-1996 #sequence_revision 05-May-1995 #text_change 04-Feb-2000

C:Accession: A27720; A0643

R:Lax, I.; Johnson, A.; Howk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Vennstr

Mol. Cell. Biol. 8, 1970-1978, 1988

A:Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mou

A:Reference number: A27720; MUID:88261272; PMID:3260329

QY 423 SVFONLQVIRGILHNGAYSLTLOGLGTSMLGLRSLRELGLALIHNNTHLCFVHTVPW 482
DB 418 HAFENLEIIRGTRKQHQFSLAVVGLNITSGLRSLKEISDDVDVIISGRNRLCVANTINW 477
QY 483 DQLFRNPHQALLHTANPEDECVGEGGLACHQLCARGHCWGPGTQCVCNCSOFLRGQECVE 542
DB 478 KKLFGTNPQTKIWNNAEKDKAVNHCNPLCSESGCWGPEPRDCVSCQNVSRGECVE 537
QY 543 ECRVLOGLPREYNARHCLPCHPCQONGSVTCFGEADOCVACAHYKDPFCVAPRCS 602
DB 538 KCNILEGEPRFVENSECICQHPCLPQAMNITCTGRPDNCICQCAHYIDPHCVKTCPA 597
QY 603 GVKPDLSPMTWKPDEEGACQPCINCTHSCVDLDDKGCPEAQSRASPLTSIIVSAVVGIL 662
DB 598 GIMGENNTL-VKTYADANNVCHLCHANCTYCAGPGLQCEVWPSGPKIPISIATIGUGL 656
QY 663 L---VVVLGVVFGILIQIY---KANSKFIGITEL---PLTFSGAMPNQAOMRILKETELRK 714
DB 657 LFIIVVVALGIGLFMRRRHIVRKRTLRLQLQERLEVEPLTPSGEAPNQAHLRLKETEPKK 716
QY 715 VKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRENTSPKANKIILDEAYVWAGVSPYV 774
DB 717 IKVLGSGAFGTYYKGLWIPEGEKVKIPVAIKELREATSPKANKIILDEAYVWASVDNPHV 776
QY 775 SRLGIGICTSTVOLAVTOLMPYGCLLDHVRENHGRGLGSDLLNWCWQIAKGMYSILEDVRLV 834
DB 777 CRLGICUTSTVQITQIMPYGCCLLDYVREHKNIGSOYLLNWCVQIAKGMYSILEDVRLV 836
QY 835 HRDLAARNVLKSPNHVKITDFGLARLLDDIDETEHADGGKVPKIMMALESILRRRPTHQ 894
DB 837 HRDLAARNVLKTPQHVKITDFGLAKLGAEEKEYHAEGKVPKIMMALESILHRIYTHQ 896
QY 895 SDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITDVMIMVKCMID 954
DB 897 SDVMSYGVTVWELMTFGSKPYDGIIPASDISILEKGERLPQPPICITDVMIMVKCMID 956
QY 955 SECPREPRELVSERWARDPQRFVLIQ-NEDLGPASPLDSTFVRSLELDDMDGLVDAAE 1013
DB 957 ADSRPFRELILEFSQWARDPQRFVLIQGDHMLPSTDSNFYRALMDEEDMEDVDAD 1016
QY 1014 EYLVPOQGFPCDPAPGAGGMVHRHRSSTSRSGGDLTLGLEPSEEAAPRSPLAPSEGA 1073
DB 1017 EYLVPOQGF-----NSPST-----SRTPLLSLSA 1042
QY 1074 GSDVFDGDLGMAKGLQSLTPHDPSPLOQRYSEPTVPLPSET--DGVPALTCSPQPEY 1131
DB 1043 TSN-----NSTVACINRNGSCRVKEDAFLOQYSSDPTGAVTEDNIDDAFL-----PVPEY 1092
QY 1132 VNQPDVRPQPSREGPLPAARPAAGATLERAKTLSPGKNGVVDVFAFGGAVENPEYL-T 1190
DB 1093 VNQ-SVPRPAGSVQNPVYHNPQHP-----APGRDLHYQN--PHSNAGVPEYLN 1141
QY 1191 PQGGAAPQPHPPPAFSPAFNLYYWQ-----DP-----PERGAPPTFKGTPT 1234
DB 1142 AQ-----PTCLSSGFSNPAWIKGSHQMSLDNPDYQDFFPKETKNGIFKG-PT 1191
QY 1235 AENPEYLGLDVP 1246
DB 1192 AENAEYLRVAPP 1203

A:Accession: A27720
A:Molecule type: mRNA
A:Residues: 1-1223 <LAX>
A:Cross-references: GB:M20386
R.Nielsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1985
A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A:Reference number: A00643; MUID:85228222; PMID:2988784
A:Accession: A00643
A:Molecule type: mRNA
A:Residues: 585-1223 <NIL>
A:Cross-references: GB:M10066
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
specific protein kinase
F:1-30/Domain: signal sequence #status predicted <SIG>
F:31-1223/Product: epidermal growth factor receptor #status predicted <MAT>
F:31-654/Domain: extracellular #status predicted <EXT>
F:31-307/Domain: EGF receptor extracellular domain repeat <EE1>
F:397-610/Domain: EGF receptor extracellular domain repeat <EE2>
F:655-677/Domain: transmembrane #status predicted <TM>
F:678-1223/Domain: intracellular #status predicted <INT>
F:719-984/Domain: protein kinase homology <KIN>
F:727-735/Region: protein kinase ATP-binding motif
F:136,202,280,361,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent) #
F:192,650/Binding site: carbohydrate (Ser) (covalent) #status predicted
F:687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:754/Active site: Lys #status predicted
F:1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat

Query Match 44.4%; Score 3011.5; DB 1; Length 1223;
Best Local Similarity 47.8%; Pred. No. 4.5e-117;
Matches 620; Conservative 175; Mismatches 349; Indels 153; Gaps 29;

Qy 8 RWGLLLALLPGAA-----STVCTGDMKRLRPASPETHLDMRLHYQQCVVQGNLE 61
Db 13 RGAAVLVLLGLGVALCSAVEKKVCCQGTNNKLTQLGHVEDHFTSLQRMYNCEVVLNLE 72
Qy 62 LTYLPTNASLFLQDIQVQGVLIQVQVPLQRLRIVRGTQLPEDNYALAVLNGD 121
Db 73 ITYEHNRLTFLKTIQVAGVLIALNMDVIPLENIQIRGNVLYDNSFALAVLSYH 132
Qy 122 PLNNTPTVTGASPGGLRELQRLSLTEILKGVLIQRLNPQLCYQDTILWKDIFHKNNOIAL 181
Db 133 -MNKTQ-----GLRELPMKRLSEILNGVKISNNPKLCNMOTVLWNDDIIDSRLK-PL 182
Qy 182 TLID-TNRSRACHPCSPCKGSRGWSESSEDCQSLTRTVACGGCA-RCKGPLPTDCCHQ 239
Db 183 TVLDPASNLSCPCKHPNCTEDHCWGAGEONCQTLTKVICAQCSGRCGRKVPDCCCHQ 242
Qy 240 CAAGCTGPKHSDCLACHFNHSGICELHCPALVTNTDTFESMNPGRYTFGASCYTAC 299
Db 243 CAAGCTGPRESDCLACRFRDATTCKTCCPLVLYNPTYQMDVNPCKYSGFATCVREC 302
Qy 300 PYNLYSTDVGSCTVLCPLHNOEVTAEADGTQRCCKSKPCARVCYGLGQMYIKANSKFTGI 359
Db 303 PHNVVTVHDGSCVRSNTDTEYV-EENGVRCKCKDGLCSKVCNGIGELKGLIS-INA 360
Qy 360 TELE-FACKKIPGLAPLPSFGDPDASNTAPLOPEOLOVFETLEETGYLYTSAMPDS 418
Db 361 TNIDSFKNCTKINGDVSILPVAFLGDAFTKTLPLDPKKLDVFTVKEISGFLLIQAMPDN 420
Qy 419 LPDLVSFONLQVIRGILHNAGYSITLQGLISWGLSRLSGLALIHNTLFCVH 478
Db 421 ATDLVAFENLEIRTKHQGYSLVAVNLKIQSLGLSLKEISDGDITAMKNKLCYAD 480
Qy 479 TVPMDQLFRNPQALLHTANPEBCVGEGLACHOLCARGHCWGPPTQCVNCSQFLRGQ 538
Db 481 TMNWRSLFATOSQTKIILQNRKNDCTADRHVCVPLCDVCGWGPFGPHCFSCFFSRQK 540
Qy 539 ECVEECRVLOGLPREYVNAHCLPCHPECPQNG---SVTCFGPEADQCVACAHYKOPPF 595

Db 541 ECVKQCNILQGE3PREFERDSKCLPCHSECLVQNSTAYNTTCSGGPGPHCMKCAHFIDGPH 600
Qy 596 CVARPCSGVKPDLISYMIWKFPPDEBEGACQPCPINCTHSCVDLDDKGCPCAEQASPLTSIV 655
Db 601 CVKACPAVLGENDTL-VMKYADANAVCOLCHPNCTRGCKPGLEGCP---NGSKTPIA 656
Qy 656 SAVV-GILLVVVLGVFGIILI--QYI---KANSKFIIGITEL--PLTPSGAMPNQAQWIRL 707
Db 657 AGVVGGLLCLVVGVLGIGLYLRRRHIVRKRTLRLQLQERELVEPLTPSGEAPNQAHLRL 716
Qy 708 KETELRKVKVLSSGAFCTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYMA 767
Db 717 KETEFKKVKVLSSGAFCTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYMA 776
Qy 768 GVSGPYVSRLLIGLICLTSTVOLVLTQLMYPGCLLDHVRNRLGSLQSDLLNWCMIQAKMSY 827
Db 777 SVDPNPHVCRLLIGLICLTSTVOLVLTQLMYPGCLLDYIREHKONIGSQYLLNWCMIQAKMSY 836
Qy 828 LEDVRLVHRDLAARNVLKSPNHVKITDFGLARLIDIDETEHADGGKVPKIMWMALESIL 887
Db 837 LEERLVRDLAARNVLKTPQHVKITDFGLAKLGADEKEYHAEGCKVPKIMWMALESIL 896
Qy 888 RRRPTHQSDVMSYGVTVWELMTFGAKYDGIIPAREIPDLLEKGBRLPOPPITCTIDVYIM 947
Db 897 HRIYTHQSDVMSYGVTVWELMTFGSKPYDGIIPAREIPDLLEKGBRLPOPPITCTIDVYIM 956
Qy 948 VKCMIDSECRPRFRELVSFSSRMARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDDDM 1006
Db 957 VKCMIDADSRRKRELIAFESKWARDPPRYLVIQDGMHLPSTDSKYRTLMBEDM 1016
Qy 1007 GDLVDAEYLVPOQGFPCPDPAFCAGGMVHRRHSSSTRSGGGDLTLGLEPSEEEPRSP 1066
Db 1017 EDIVDAEYLVPHQGF-----NSPST-----SRTPL 1042
Qy 1067 L-----APSGASGVDPDGLGMGAAGLQSLPDRSPQRYSEDDTVLPST--DGY 1119
Db 1043 LLSLSLATSNNATNCID-----RNGQGHVPREDSFVQRYSSDPTGNFLEESIDDF 1094
Qy 1120 VAPLTCSPQEPQNPQPPSPREGPLPAARPAAGATLERAKTLPSPGKGVVVDYF-- 1177
Db 1095 L-----PAPEYVNO--LMPKKFS-----TAMVQNIYNNISLT 1125
Qy 1178 -----AFGAVENPEYLTPOGGAAPQHPHPPAFSPAFDNLVYWDQ----- 1217
Db 1126 AISKLPMDSRYQNSHSTAVDNPEYL-----NTNQSLAKTVFESSPYWIOQSNHOIN 1177
Qy 1218 -DPPE-----RGAPSTFKGTPTAENPEYLGLDVP 1246
Db 1178 LDNPDYQDFLNPETKPNGLLKVPAENPEYLRVAAP 1214
RESULT 7
A:Accession: A47253
epidermal growth factor receptor, HER4 - human
C:Species: Homo sapiens (man)
C:Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C:Accession: A47253
R:Piowman, G.D.; Culouscou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.; Ne
Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993
A:Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epidermal
A:Reference number: A47253; MUID:93189574; PMID:8383326
A:Accession: A47253
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-1308 <PLO>
A:Cross-references: GB:L07868; NID:G337359; PIDN:AAB59446.1; PID:G337360
A>Note: sequence extracted from NCBI backbone (NCBIP:126842)
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; growth factor receptor
F:716-981/Domain: protein kinase homology <KIN>
F:724-732/Region: protein kinase ATP-binding motif

[illegible][illegible]

```
Db 123 YQK-NPSSP--DYYQVGLKQLQLSNLTELISGGVVKVSHNPLNCLLVETINWMDIVDKTNP 179
Qy 180 ALTLDTNRSRACHPCSPMKSGRCWGESSEDQSLTRTVACAGC-ARCKGLPLPTDCHE 238
Db 180 TWNLIHAFERQCKQCDHGVNCSWAQPGHCKQKFTKLLCAEQCNRRCRGPKPIDCCNE 239
Qy 239 OCNAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCTA 298
Db 240 HCAGCTGPRATDCLACRDFNDGCTCKDTPPKIYDIVSHQVNDPNIKYTFGAACVKE 299
Qy 299 CPYNLSLTVGSGTLCVPLHNOETVAEDGTORCEKSCPKARCVVGLGMQVLIKANSKFIG 358
Db 300 CPSNVVTE-GACVRSKCSAGMLEVD-ENGKSKCPDGVCPKVDGIGIGSL-SNTIAVN 356
Qy 359 ITEL-ERAGCKIKFISLAFLPESFDGDPASNTAPLOEQLOVETLEBITGYLISAWPD 417
Db 357 STNIRSFNCTKINGDIIILNRNSPEGDPHYKIGTMDPEHLNLTITVKEITGYLVIMWPE 416
Qy 418 SLPLSVFQNLQVIRGRILHNGAYS-LTLQGLGISWGLRSLRELGLALIHNTLHLCF 476
Db 417 NMTSLSVFQNLLEIRGRTTFRSGFSFVVVQVRHLQWLGLRSLKEVSAGNVILKNTLQRY 476
Qy 477 VHTVPWDLFRPHOALLHTANRPEDECVGEGILACHQLCARGHCWGPCTOCVNCQFLR 536
Db 477 ANTINWRFLFRSDOSIEYDART-----ENQTCNNECEDGCW-PGPTMCVSCLVHVR 528
Qy 537 QOECVECRVLQGLPREYNARHCLPCHPEQOPQNGSVTCFGEADQCACAHYKDDPPFC 596
Db 529 GGRVCASNLQGBEAREAQVGRGVQCHOECLVQDLSLTCYGPANCKSAHFQDGPQC 588
Qy 597 VARCPGKVPDLSPYPIWKFPDEBAGCQCPINCTHSCVDLDDKCPAEQASPLTSIVS 656
Db 589 IPRCPHGILGDDTL-IWKYADKMGCCQPCQNCCTQGCSPGLSGCRGD-IVSHSLAVG 646
Qy 657 AVVGLLNVVLGVFGILIQIKANSK----FIGITEL--PLTPSGAMPNQAQRILKE 709
Db 647 LVSGLLITIVALLVILRRRRIRKRTIKCLQEKELVBLPLTPSGQAPNQAFLRILKE 706
Qy 710 TELRKVKLGSGAFGTVYKGIWIPDGENVKIPVAKVLRNTSPKANKEIIDEAYVWAGV 769
Db 707 TEFKKDRVLGSGAGFTVYKGLWNPGENIRIPVAKVLRATSPKVNQEVLEDEAYVWASV 766
Qy 770 GSPVYSRLGLCLSTVOLVTQLMPYGCILLDHHVRENRRGLSGQDLLNWCMIQAKMSYLE 829
Db 767 DHPHVCRLLGLCLTSAVOLVQLMPYGCILLDVVRQHERICQWLLNWCVOIAGKMNYLE 826
Qy 830 DVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEXHADGKVPKIMMALESILRR 889
Db 827 ERHLVHRDLAARNVLLKNPNHVKITDFGLSKLLTADKEBYQADGGKVPKIMMALESILQW 886
Qy 890 RFTHQSDVWSYGVTVWELMTFCAPYDGIIPAREIPDLLEKGERLPQPPICITDVYTMIVK 949
Db 887 TYTHQSDVWSYGVTVWELMTFGSKPYDGIIPAKETASVLENERLPQPPICITIEVMIILK 946
Qy 950 CWMIDSECRPRELIVSEFSRMARDPQRFVVIQNEEDLGPASPLDSTFYRSLLDDMDGDL 1009
Db 947 CWMIDPSSRPREFRELIVGFSGQWADPSRYLVIQ---NLPSLDRRLFSRLSSDD--DV 1001
Qy 1010 VDAEYLVPQOGFPCPDAPAGGGMVHRHRSSTRSGGDLTLGLEPSEEAAPRSLAP 1069
Db 1002 VDAEYLLPLPKRI-----NRQGS-----EPCIP 1024
Qy 1070 SEGAGSDVDFDGLGMAKGLQSLTPHDPSPLOQRYSEDPV-PLPSETDGVVAPLTCSPQ 1128
Db 1025 PTGH-----PVRENSITLRNISDPTQNALKEKLDLGH----- 1055
Qy 1129 PEYNQPDVRPOP-----PSPRE-----GELP-AARPAGATLERAKTLSPGRKNGVVKD 1175
Db 1056 -EYVNPQSGTSSRLSDIYNPNYEDLTDCWGPVSLSSQEAETNFSRPEYLNTQNSL--- 1111
Qy 1176 VFAFGGAVENPEYLTPOGGAAPQHPHPAPSPADNLYWQDDPPRGAPSTPKGTPTA 1235
Db 1112 PLVSSGSGMDDPDY---QAG-----YQAAF-----LPQTGALTGMWFLPAA 1149
```

```
Qy 1236 ENPEYLK 1242
Db 1150 ENLEYLG 1156
```

RESULT 9

A36223

kinase-related transforming protein (erbb3) (EC 2.7.1.1-) precursor - human

C:Species: Homo sapiens (man)

C>Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000

C:Accession: A36223; I59164

R:Kraus, M.H.; Issing, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.

Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989

A:Title: Isolation and characterization of ERBB3, a third member of the ERBB/epidermal g

A:Reference number: A36223; MUID:90083234; PMID:2687875

A:Accession: A36223

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1342 <KRA>

A:Cross-references: GB:M29366

R:Plowman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todaro, G.J

Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990

A:Title: Molecular cloning and expression of another epidermal growth factor receptor-re

A:Reference number: I59164; MUID:90311312; PMID:2164210

A:Accession: I59164

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-559, 'G', 561-957, 'F', 959-1063, 'G', 1065-1342 <RES>

A:Cross-references: GB:M34309; NID:G183990; PIDN:AAA35979.1; PID:G306841

C:Genetics:

A:Gene: GDB:ERBB3; HER3

A:Cross-references: GDB:119880; OMIM:190151

A:Map position: 12q13-12q13

C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo

C:Keywords: ATP; phosphotransferase

F:707-972/Domain: protein kinase homology <KIN>

F:715-723/Region: protein kinase ATP-binding motif

Query Match

35.1%; Score 2381.5; DB 2; Length 1342;

Best Local Similarity 40.1%; Pred. No. 4.5e-91;

Matches 526; Conservative 197; Mismatches 450; Indels 139; Gaps 35;

```
Qy 10 GLLIALLPPGAA--STQVCTGTDMKRLRPLASPETHLDMRLHYQGCQVQVQGNLELTPLT 67
Db 11 GLUFSLARGSEVGNQAVCFSTLNGLSVTGDAENQYQTLKLYERCEVVMGNLIVLTGH 70
Qy 68 NASLSFLQDIQEOGVYVLIHNVQVQVPLQRLRIVRGTFEDNYALAVLDNGDPLNNTT 127
Db 71 NADLSFLQWIREVTGYVLVAMNPFSTLPLNLRVVRGTQVVDGKFAIVM----LNVNT 125
Qy 128 PVTGASPGGLREQLRSLRTEILKGGVLIQRNPOLCYQDTILWKDIFHKNQLALTLDTN 187
Db 126 ----NSSHALRQLRLTQLTEILSGGVYIERNDKLCHMDTIDWRDIVDRD---AEIVVKD 178
Qy 188 RSRACHPCSPMKSGRCWGESSEDQSLTRTVACAGC-ARCKGLPLPTDCHEQCAAGCTG 246
Db 179 NGRSCPCHEVCVKG-RCWGPGSEDDCQLTKTIICAPQCNHGFCHGFNPNQCHDCAGCGSG 237
Qy 247 PKHSDDLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNYLST 306
Db 238 PQDTDCFACRHFNDGACVPRCPQPLVYNKLTLEENPHTKYQYGGVGVASCAPHNFV-V 296
Qy 307 DVGSCSTLVCLPHNQEVTAEDGTORCEKSCPKARCVVGLGMQVLIKANSKF--IGITEL- 363
Db 297 DQTSQVRAACPDKMEVD-KNGLKMCPCGGLCPKACEGTG-----SGSRFTQVDSNNIDG 350
Qy 364 FAGCKIKFISLAFLPESFDGDPASNTAPLOEQLOVETLEBITGYLISAWPDSLPLDLS 423
Db 351 FVNCTKILGNLDFLITGLNGDPWHKIPALDPEKLVNPRVREITGYLNIOSWPHHMFNS 410
Qy 424 VFQNLQVIRGRILHNGAYS-LTLQGLGISWGLRSLRELGLALIHNTLHNTLHLCFVHTVPW 482
```


Db 761 VGSLDHAIHVRLLGLCPGSSIQLVQYLPGLSLLDHVHQHRTLGPQLLLNKGVOIAKGM 820
QY 826 SYLEDVRLVHRDLAARNLVKSPNHNKVTDFGLARLIDIDETEHADGKVPKIMMALES 885
Db 821 YLLEHSMVHRDLARNVWLKSPSQVQVADLLPPDDKQLHSEAKTPIKMALES 880
QY 886 ILRRFTTHQSDVWSYGVTVWELMTFGAKPYDGIIPARETIPDLLEKGERLPQPPICITIDVYM 945
Db 881 IHFGKYTHQSDVWSYGVTVWELMTFGABPYAGRLAEITPDLLEKGERLAQPOICTIDVYM 940
QY 946 IMVKCWMIDSECRPRFRELVSFSESMARDPQRFVVIQNEGLCPASPLDSTFYRSLLLEDD 1005
Db 941 VNVKCMIDENIRPTFKELANFTKWARDPPRYLVIKRAS-GPGTP--PAEPPSVLTITKE 997
QY 1006 MGDLDVAEYLVPOQGFCCPDPAFGAGGMVHRRHSSSTRSGGDLTLGLPESBE----- 1060
Db 998 L-----QEALEPEL-----DLDLLEABEGLATS 1023
QY 1061 -----EAPRSLAPSEG-----AGSDVFDGLGMAAGLQSLPHTD 1097
Db 1024 LGSALSPTGLTRPRGQSLLSPSSGYMPNMNQSSLGAECLDSAVLGGREGQFSRPSILH- 1082
QY 1098 PSLQRYSEDPTVPLPSTGDGVV---APL-----TC-----SPOPE-----YVNOPDV 1137
Db 1083 PIPGR-----PASESEGHVTGSEAEALQEKVSVCRSRSRSPRPGDSAYHSORHS 1135
QY 1138 RPQPPSPREGP-----LPAARPAAGATLERAKTLSP-GKNGVV-----KDVPAF 1179
Db 1136 LLTPVPLSPGLEBEDGNGVMPDTHLGRASSREGTSSVGLSSVLTGTEDEED----- 1191
QY 1180 GGAVENPYLTQGGAAAPQPHPP 1202
Db 1192 ----EVEYMNKRKRGSP-PRPP 1209

RESULT 11
TVFVLV
N:Contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase
C:Species: avian leukosis virus, ALV
C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 11-Jun-1999
C:Accession: B00643; A00643
R:Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1985
A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and processing of the gag-env-erbB fusion protein
A:Reference number: A00643; MUID:85228222; PMID:2988784
A:Accession: B00643
A:Molecule type: mRNA
A:Residues: 1-698 <Nil>
A:Cross-references: GB:M10066; GB:M13881; NID:g211749; PID:AAA48763.1; PID:g211750
A:Note: in Genbank entry CHKERBBF, release 109.0, the source is designated as Gallus gallus
C:Comment: This protein is synthesized as a gag-env-erbB protein.
C:Genetics:
A:Gene: gag-env-erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific protein kinase
F:1-6/Product: gag protein (fragment) #status predicted <GAG>
F:7-59/Product: env protein (fragment) #status predicted <ENV>
F:60-698/Product: protein-tyrosine kinase envB #status predicted <ERB>
F:194-459/Domain: protein kinase homology <KIN>
F:202-210/Region: protein kinase ATP-binding motif
F:229/Active site: Lys #status predicted

Query Match 24.8%; Score 1683.5; DB 1; Length 698;
Best Local Similarity 50.6%; Pred. No. 1.4e-62;
Matches 362; Conservative 82; Mismatches 140; Indels 131; Gaps 20;

QY 578 GPEADQCACAHYKDPPEPCVRCFSGVKPFDLSYMPKWPFBEGACQPCNCTHSCVDL 637
Db 60 GP--DHCWKCAHFDGPHCVKACAGVLGENDTL-VWKYADANAVCOLCHNCTRGCKGP 116
QY 638 DDKGCAPAEQRASPLTISVSAYV-GILLVVLGVVFGILI--QYI---KANSKFIITEL- 690

Db 117 GLECP--NGSKTPSIAAGVVGGLCLLVVVGIGLYLRRRHIVRKTRRLRLQERELV 173
QY 691 -PLTPSGAMPNQAOMRIILKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAKVLRE 749
Db 174 EPLTPSGEAPNQAHLRIILKETEFKKVKVLGSGAGFTVYKGLWIPGEKVKIPVAKELRE 233
QY 750 NTSKANKELDEAYVWAGVSPVSRLLGICLTSTVOLVTQLMPIYCGCLLDHVRNRRGL 809
Db 234 ATSPKANKELDEAYVWASVDNPHVCRLLGICLTSTVQLITQLMPIYCGCLLDVIREKONI 293
QY 810 GSQDLLNKCWQIAKMSYLEDVRLVHRDLAARNVLKSPNHNKVTDFGLARLIDIDETEH 869
Db 294 GSQYLLNKCWQIAKGMNVLERRLVRDLAARNVLVKTTPQHKITDFGLAKLGADEXEY 353
QY 870 HADGKVPKIMMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPARETIPDLLEK 929
Db 354 HAEGKVPKIMMALESILHRYTHQSDVWSYGVTVWELMTFGSKPYDGIIPASEISSVLEK 413
QY 930 GERLPQPPICITIDVYIMVKCWMIDSECRPRFRELVSFSESMARDPQRFVVIQ-NEDLGP 988
Db 414 GERLPQPPICITIDVYIMVKCWMIDADSRPKFRELIAEFSKWARDPPRYLVIQDERMHL 473
QY 989 ASPLDSTFYRSLLLEDDMDGLVDAEYLVPOQGFCCPDPAFGAGGMVHRRHSSSTRSGG 1048
Db 474 PSPTDSKPYRTLMEEDEMDIVDAEYLVPHQGF-----NSPST----- 513
QY 1049 GDLTLGLEPSEEBEAPRPL-----APSEGAGSDVFDGDLGMAAGLQSLPHTDPSPLQR 1103
Db 514 -----SRPLSSLSATSNNSATNCID-----RNGQGHVPRDSFVQR 551
QY 1104 YSEDPTVLPFSET--DGYPVAPLTCSPQPEYVQPDVPRQPPSPREGPLPAARPAAGATLER 1161
Db 552 YSSDPTGNFLEESIDGFL-----PAPEYVQ--LMPKPS----- 585
QY 1162 AKTLSPGKNGVVKDVF-----AFGAVENPEYLTPOGGAAPQPHPPPAFS 1206
Db 586 ---TAMVQNOIYNNISLTAISKLPMSRYONSHSTAVDNPEYL-----NTNQSPLAK 634
QY 1207 PAFDNLVYWDQ-----DPPE-----RGAPPSTFKGTPTAENPEYLGLDVP 1246
Db 635 TVFESSYWIQSGNHQINLNDPVDYQDFLNETKPNGLLKVPAENPEYLRVAAP 689

RESULT 12
TVYUHV
N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain H)
C:Species: avian erythroblastosis virus
C:Date: 18-Apr-1994 #sequence_revision 18-Apr-1984 #text_change 11-Jun-1999
C:Accession: A00644; A38022
R:Yamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.
Cell 35, 71-78, 1983
A:Title: The erbB gene of avian erythroblastosis virus is a member of the src gene family
A:Reference number: A00644; MUID:84026539; PMID:6313229
A:Accession: A00644
A:Molecule type: DNA
A:Residues: 1-604 <YAM>
A:Cross-references: GB:X01216; NID:g209676; PID:AAA42400.1; PID:g209678
R:Debuire, B.; Henry, C.; Benaissa, M.; Biserte, G.; Claverie, J.M.; Saule, S.; Martin, Science 224, 1456-1459, 1984
A:Title: Sequencing the erbA gene of avian erythroblastosis virus reveals a new type of protein kinase
A:Reference number: A38022; MUID:84223957; PMID:6328658
A:Accession: A38022
A:Molecule type: DNA
A:Residues: 1-28, 'W', 30-139, 'F', 141-145, 'V', 147-152 <DEB>
A:Cross-references: GB:X02006
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific protein kinase
F:130-395/Domain: protein kinase homology <KIN>
F:138-146/Region: protein kinase ATP-binding motif
F:165/Active site: Lys #status predicted

Query Match 23.9%; Score 1620; DB 1; Length 604;
Best Local Similarity 50.5%; Pred.No. 4.9e-60;
Matches 348; Conservative 78; Mismatches 131; Indels 132; Gaps 18;

Qy 587 CAHYKDPFFCVARCPGSKPDLSPMTLWKPDEEGACOPCINCTHSCVDLDKKGPAEQ 646
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3 CAHFIDGPHCKVAKAGVLGENTL-VRKTYADANAVCOLCHPNCTRCKGFGUEGCP--- 58
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 647 RASPLTSIVSAVV-GILLVVVLGVVFGLI--QYI---KANSKFIGITEL--PLTSGAM 698
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 59 NGSKTFSIAAGVVGGLCLTVVGLIGLYLRHHVRKRTLRLQLQERELVELPTSGEA 118
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 699 PNOAQRILKETELRKKVLGSGAFGYVYGIMPGENVKIPVAIKVLENTSPKANKE 758
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 119 PNOAHLRIKETBFKKVKVLGSGAFGIYKGLWIPEGEKVKIPIAIKELREATSPKANKE 178
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 759 ILDEAYMAAGVGPYSYRLGICLTSTVOLATOLMPYGCLLDHVRENRGFLGSODLLNWC 818
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 179 ILDEAYMASVDNPHVCRLGICLTSTVQITQLMPYGCLLDIREHKONIGSQYLNNWC 238
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 819 MQIAGMSYLEDLVRLHRLDAARNLVKS PNHVKITDFGLARLLDIDETEHADGGKVI 878
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 239 VQIAGMNVLEERRLVHRDLAARNLVKTPOHVKITDFGLAKLLGADEKEYHAEGKVI 298
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 879 KWALESILRRFTHQSDVMVS YGVTWELMTFGAKPYDGIPAREIPDLLEKGERLPQPI 938
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 299 KWALESILHRIVTHQSDVMVS YGVTWELMTFGSKPYDGIPASEISSVLEKGERLPQPI 358
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 939 CTIDVYMIMVKCMWDISECPRELVSERSWARDPORFVIO-NEDLGPA SPLSTFFY 997
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 359 CTIDVYMIMVKCMWDADSRPKRELTAESKWARDPPRYLVIQDERMELPSPTDSKFY 418
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 998 RSILLEDDMDGLVDAEYLVPOQFFCPDPAPGAGGMVHRHSRSSSTRSCGGDLTLGLEP 1057
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 419 RTLWEEDMEDIVIDADYLVPHQGFF-----NSPST----- 449
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1058 SEEEAPRSPL-----APSEGA GDVFDGD LGMAAKGLSLTHDSP LQRYSEDPTVPL 1112
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 450 -----SRTPLLSSLSATSNN SATNCID-----RNQGHVPVREDSFVQRYSSDPTGNF 496
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1113 PSET-DGYVAPLTCSPQYVYNQDVPRPOPSPREGPLPAARPAGATLERAKTLP SGKKN 1170
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 497 LEESIDGFL-----PAPEYVNQ--LMPKKPSTAM----- 524
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1171 GVWKDVFAF-----GGAVENPEYLT PQGGAA POPHPFP AFSPA FN 1211
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 525 -VNQIYFNFLS LAISKLPMSRYQN SHSTAVDNPEYL-----NTNOSP LAKTVFES 575
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1212 LYTWDDQPBERGAPPSTFKGTPTAENPEY 1240
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 576 SPYWIQSNGHQ-----INLDNPDY 594
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 13

GQFFE
epidermal growth factor receptor - fruit fly (*Drosophila melanogaster*)
N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
C:Species: *Drosophila melanogaster*
C>Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 11-Jun-1999
C/Accession: A00640; A38021
R:Livneh, B.; Glazer, L.; Segal, D.; Schlessinger, J.; Shilo, B.Z.
Cell 40, 599-607, 1995
A>Title: The *Drosophila* EGF receptor gene homolog: conservation of both hormone binding
A/Reference number: A00640; MUID:85124611; PMID:2982499
A/Accession: A00640
A/Molecule type: DNA
A/Residues: 1-1330 <LIV>
A/Cross-references: EMBL:K03054
R:Wadsworth, S.C.; Vincent III, W.S.; Bilodeau-Wentworth, D.
Nature 314, 178-180, 1995
A>Title: A *Drosophila* genomic sequence with homology to human epidermal growth factor re
A/Reference number: A38021; MUID:85137938; PMID:2983232
A/Accession: A38021

A:Molecule type: DNA
A:Residues: 'A',832-866,'V',868-943,'QTPSLVK' <WAD>
A:Cross-references: EMBL:X03293; NID:G7922; PIDN:CAA26157.1; PID:G929565
C:Comment: This sequence is tentative because the introns have not been identified.
C:Genetics:
A:Gene: FlyBase:Egfr
A:Cross-references: FlyBase:FBgn0003731
A:Map position: 2 57F
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
F:1-732/Domain: extracellular #status predicted <EXT>
F:733-764/Domain: transmembrane #status predicted <TM>
F:765-1330/Domain: intracellular #status predicted <INT>
F:808-1072/Domain: protein kinase homology <KIN>
F:816-824/Region: protein kinase ATP-binding motif
F:122,300,324,363,518,688,695,700/Binding site: carbohydrate (Aen) (covalent) #status pr
F:774/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:843/Active site: Lys #status predicted
F:1181/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status predict

	Query Match	23.8%	Score 1612.5;	DB 1;	Length 1330;	
	Best Local Similarity	29.6%	Pred. No. 2.1e-59;			
	Matches 409;	Conservative 175;	Mismatches 413;	Indels 385;	Gaps 39;	
Qy	80	VOGYVLI	AAHNOVRQVPLQR	LRIVRTQLF-----	EDNYALAVLDNGDPLNNTPTVTGASP 134	
Db	38	ITNYIVIGLD	IPCTLSYRLQIR	TLTIRTLFSLSV	EEKYALFV-----TY 81	
Qy	135	GLRELQ	LRSLTEILKGGV	LQNRNPQLCVQD	TLWKDI FHKNNQLALTLDITNRSRACHP 194	
Db	82	SKMYTLEIP	DRDLVNLGQVGFHNN	YLNCHMRTIQWSEIV	NSGTDAYNYDYFTAPERSCP 141	
Qy	195	CSPMCKSG	RCWGESSEDQSL	TRTVCCAGCA--	RCKGPLPTDCCHQCQAAGCTGPKHSDC 252	
Db	142	CHESCTHG-	CWGEGPNCQKPS	KLTCSPQCAGRC	CVGPKPRECHLFCAGCGTGPTQKDC 200	
Qy	253	LACLHFNH	SGICELHCPALV	TYNTDTFESMP	NPEGTYTFGASCVTACPNYNLSTDVGSCT 312	
Db	201	IACKNFDE	AVSKESCEP	PMRKYNPTTYVLE	TNPFGKYAYGATCVKECP-GHLLRDNGACV 259	
Qy	313	LVCPLHQE	VTAE	DTQRCCKSKPCAR	VCYGLGMQYIKANSKFIGITEL-----EFAG 366	
Db	260	RSCPQDKM	DKGGE-----	CVPCNGCPKTC-----	PGVTVLHAGNIDSPRN 300	
Qy	367	CKKTFGSL	AFLPESPDG--	DPASNTA-----	PLQPEQLQVFETLBEITCYLIVISAWPDS 418	
Db	301	CTVIDGNR	ILDQTFSGDQV	VYANTMGPRYIPLD	PERREVFSTVKEITCYLALIEGTHPQ 360	
Qy	419	LPDLSVFQ	NQVLRILHNGAY-	SLTLQGLGISW	GLRSLRSLRELGSGLALIHNNTHLCFV 477	
Db	361	FRNLSPRN	LETIHGRQLMES	FAALAI VKSSLY	SLEMRNLKQISGGSVVVIQHNRLCYV 420	
Qy	478	HTVPWDQ	LFRPHQALLHTAN	RPEDC-----	----- 504	
Db	421	SNIRWPAI	QKEPEQKVWV	NNENLRADLCGK	FLTILISVQHNIIMHIFAICREKNHLLGSV 480	
Qy	505	-----	-----	-----	----- 504	
Db	481	QRGLLGSH	GSVPYLBQLQFQW	HLHRRLLWLYIQV	SINSTDKSNHEQLTDACYSPSVPT 540	
Qy	505	-----	-----	-----	-----VG 506	
Db	541	SLTIERAR	YAIQSLAGLAME	LEQITARSASMRH	SKTLPAGRQVPRWVFLGVCASARAGIA 600	
Qy	507	EGLA-----	CHQI	CARGHCWGPGPTQ	CVNCISOFLRGCEVCECRVLOGLPREYV---N 556	
Db	601	EPLAGRAV	CRKCHPLCEL	TNNGYGEHQV	CKTHYKRREQCETEC-----PADHYTDBE 654	
Qy	557	ARHCLPCH	PECPQNGSV	TCFGEADQCV	CAHAHK-----DPPE-----CVARCPSG 603	
Db	655	QRECFQRH	PEC-----	NG-----	CTGPGADDDKSCRNFKLF	DANETGPNVNSWNTFNCSTKCPLE 708
Qy	604	VK-PDLSY	WPIWKFFDEE	GACQFCP	INCTHSCVDLDDKGC	PABQRASPLTISVSAVVGIL 662


```

Db      709 MRHVNQYTAIGPY-----CAASPPRSSKITANLDVN-----MFIITGAVLVP 752
Qy      663 LVVLGVVFGILIOYIKANSFIGIT-----ELPLTPSGAMPNQAMRILKETELRK 714
Db      753 TICILCVVYICROKQAKKETVMTALSGREDEPLRPSNIGANLCKLRIVKDAELRK 812
Qy      715 VKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVWAGVGSYV 774
Db      813 GGVLMGAFGRVYKGVWVPEGENVKIPVAIKELLKSTGAESSEEPLEAYIMASHEHVN 872
Qy      775 SRLIGICLTSTVOLTPQMLPVCGLLDHVNRGRIGSODLLNWCQIAKGSYLEDVPLV 834
Db      873 LKLLAVCMSSQMLLTQMLPGCLLDYVYRNRRDKTIGSKALLNWSQIAKGSYLEEKLV 932
Qy      835 HRDLAARNVLVK---SPNHVITDIFGLARLLDIDETEHADGGKVPKIKWMALESTLRRF 891
Db      933 HRDLAARNVLRLLAGEDH---DFGLAKLLSSDSNEYKAAGKVPKIKWLALECIERNVF 988
Qy      892 THQSDVWSVGVTVWELMTFGAKPYDGPAPRIPDLLEKGERLPQPPICTIDVYIMVKW 951
Db      989 TSKSDVWAFGVTIWELLTFGORPHENIPAKDIPDLIEVGLKLEQPEICSLDIYCTLLSCW 1048
Qy      952 MIDSECRFRFRELVSFSEMRDQRFVVIQNEIDLG--PASPLOSTFVRSILLEDD---DM 1006
Db      1049 HLDAAMRFTFKOLTTFVFAEFARDPGRYLAILGDKFTRLPA-----YTSQDEKDLIRKL 1101
Qy      1007 GDLVDAEYLVPOQGFCCPDAPAGAGGMVHRHRSSTRSGGDLTLGLEPSEEP--- 1063
Db      1102 APTTDGSAIAKAPDDYLQPKAALGPS-----HRTDCT-----DEMPKLN 1140
Qy      1064 ---RSLAPSGAGSDVFDG---DLGMAAKGLQSLPHTDPSPLQVRSYEDPTVPLPSETD 1117
Db      1141 RYCKDPSKNKSTGDDERDSSAREVGVGNLR-----LDLPVDED 1179
Qy      1118 GYVAPLTCSPQPEYVQNQDVRPQPPSPREGPLPAARPAGATLERAKTILSPGKNGVVKDVF 1177
Db      1180 DYLMPT-TCQPGNNNNNNV-----NPNQNNAAVGVAAAGYM-----DLI 1217
Qy      1178 AFGGAVENPEYL---TPGGGAAPOPH-----PPAPSP-APDNLYY 1214
Db      1218 GVPVSVNDPEYLLNAQTILGVGESPIQTIGIPVNGGPGTMEVKVMPGSEPTSSDHEY 1277
Qy      1215 WD 1216
Db      1278 ND 1279

RESULT 14
S35745
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus
C:Species: avian erythroblastosis virus
C:Date: 03-Mar-1994 #sequence_revision 26-May-1995 #text_change 28-Feb-1997
C:Accession: S35745
R:Vennstrom, B.
submitted to the EMBL Data Library, March 1993
A:Reference number: S35743
A:Accession: S35745
A:Molecule type: DNA
A:Residues: 1-544 <VEN>
A:Cross-references: EMBL:X12707
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP, oncogene, phosphotransferase; transforming protein; tyrosine-specific p
F:135-400/Domain: protein kinase homology <KIN>
F:170/Active site: Lys #status predicted

Query Match      23.1%; Score 1564; DB 2; Length 544;
Best Local Similarity 53.1%; Pred. No. 9e-58;
Matches 333; Conservative 72; Mismatches 124; Indels 98; Gaps 17;


```

```

Qy      578 GPEADQCVACAHYKDPFPFCVARCPGKPDLSYMPIWKFPDEBEGACQPCPINCHTHSCVDL 637
Db      1 GP--DHCWKCAHFIDGPHCVKACAPAGLGENDTL--VMKYADANAVCOLCHPCNTRGCKGP 57
Qy      638 DDGCPAEQRASPLTSTVSAVV--GILLVVLGVVFGILL--QYI---KANSKFIGITEL- 690
Db      58 GLEGP---HGSKTPSTAAGVVGGLCLLVVVGIGLYLRRRHIVRKRTLRLRLLQERELV 114
Qy      691 -PLTPSGAMPNQAMRILKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRE 749
Db      115 EPLTPSGEAPNQAHRLILKETEFKKVKVLGFGAGFTVYKGLWIPGEKVTIPVAIKELRE 174
Qy      750 NTSPKANKEILDEAYVWAGVSPVSRLLGICLTSTVOLTPQMLPVCGLLDHVNRGRRL 809
Db      175 ATSPKANKEILDEAYVWASVDNPHVCRLLGICLTSTVOLTPQMLPVCGLLDYIREHKDNI 234
Qy      810 GSODLLNWCQIAKGSYLEDVRLVHRDLAARNVLKSPNHVKITDIFGLARLLDIDETEH 869
Db      235 GSQYLLNWCQIAKGMNYSLEERHMHVRLDLAARNVLVKTPOHKITDFGLAKQGADEKEY 294
Qy      870 HADGGKVPKIKWMALESILRRRFTHQSDVWSVGVTVWELMTFGAKPYDGPAPRIPDLLEK 929
Db      295 HAEGKVPKIKWMALESILHRIYTHQSDVWSVGVTVWELMTFGSKPYDGPASEISSVLEK 354
Qy      930 GERLPQPPICTIDVYIMVKWMIDSECRFRFRELVSFSEMRDQRFVVIQ-NEDLGP 988
Db      355 GERLPQPPICTIDVYIMVKWMSDADSRPKFRELIAEFSSKQVARDPPRYLVIQDERMHL 414
Qy      989 ASPLDSTFYRSILLEDDMDGDLVDAEYLVPOQGFCCPDAPAGAGGMVHRHRSSTRSGG 1048
Db      415 PSPDTSKPYRTLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 454
Qy      1049 GDLTLGLEPSEEPAPRSP-----APSEGAGSDVFDGDLGMAAKGLQSLPHTDPSPLQR 1103
Db      455 -----SRPFLSSLSATSNNSATNCIDRNGG-----H----- 481
Qy      1104 YSEDPTVPLPSETDGYVAPLTCSPQPEYVQNQDVRPQPPSPREGPLPAARPAGAT-LERA 1162
Db      482 -----PYREDGFL-----PAPEYVQN--LMPKKPSTAMVQNIYNYISLTAISKL 524
Qy      1163 KTLSPGKNGVVKVFAFGGAVENPEYL 1189
Db      525 PIDSRYQN-----SHSTAVDNPEYL 544

RESULT 15
S00727
kinase-related transforming protein (erbB) (EC 2.7.1.1-) - avian erythroblastosis virus (
C:Species: avian erythroblastosis virus
C:Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 28-Feb-1997
C:Accession: S00727
R:Scotting, P.; Vennstrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.
Oncogene Res. 1, 265-278, 1987
A:Title: Common site of mutation in the erbB gene of avian erythroblastosis virus mutant
A:Reference number: S00727; MUID:88217326; PMID:2897102
A:Accession: S00727
A:Molecule type: DNA
A:Residues: 1-545 <SCO>
A:Cross-references: EMBL:X06943
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; phosphotransferase
F:135-400/Domain: protein kinase homology <KIN>
F:143-151/Region: protein kinase ATP-binding motif

Query Match      23.0%; Score 1557; DB 2; Length 545;
Best Local Similarity 53.1%; Pred. No. 1.8e-57;
Matches 333; Conservative 71; Mismatches 125; Indels 98; Gaps 17;


```

```

Qy      578 GPEADQCVACAHYKDPFPFCVARCPGKPDLSYMPIWKFPDEBEGACQPCPINCHTHSCVDL 637
Db      1 GP--DHCWKCAHFIDGPHCVKACAPAGLGENDTL--VMKYADANAVCOLCHPCNTRGCKGP 57
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 09:10:23 ; Search time 22.2855 Seconds
(without alignments)
5413.772 Million cell updates/sec

Title: SEQ4-695-709-12
Perfect score: 6815
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGDPVP 1255

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6640	97.4	1255	1 A24571	protein-tyrosine k
2	5843	85.7	1260	1 TVRTWU	protein-tyrosine k
3	5833.5	85.6	1254	2 I48161	p-185 precursor -
4	3067	45.0	1210	1 GQHUE	epidermal growth f
5	3038	44.6	1210	2 A53183	epidermal growth f
6	3014.5	44.2	1223	1 TVCHLV	epidermal growth f
7	2887.5	42.4	1308	2 A47253	epidermal growth f
8	2598	38.1	1166	1 S06142	protein-tyrosine k
9	2332.5	34.2	1342	2 A36223	kinase-related tra
10	2246.5	33.0	1339	2 JC4387	epidermal growth f
11	1686.5	24.7	698	1 TVFVLV	protein-tyrosine k
12	1623	23.8	604	1 TVYUHV	protein-tyrosine k
13	1575	23.1	544	2 S35745	protein-tyrosine k
14	1568	23.0	545	2 S00727	kinase-related tra
15	1565.5	23.0	1330	1 GQFFE	epidermal growth f
16	1551	22.8	540	2 B44776	protein-tyrosine k
17	1549	22.7	540	1 TVFVFB	protein-tyrosine k
18	1503	22.1	644	2 A36325	epidermal growth f
19	1243	18.2	1323	2 E88257	protein let-23 lim
20	1243	17.0	1374	2 S70712	protein-tyrosine k
21	1160	17.0	1369	2 S70713	protein-tyrosine k
22	1126	16.6	1717	1 A45538	epidermal growth f
23	1126	16.5	527	2 A42032	epidermal growth f
24	975.5	14.3	843	2 A27131	epidermal growth f
25	806.5	11.8	346	2 S13807	protein-tyrosine k
26	754.5	11.1	311	2 S13808	protein-tyrosine k
27	700	10.3	1363	2 T43220	insulin-like growth
28	666	9.8	1383	2 A36080	insulin receptor p
29	665.5	9.8	1372	2 A34157	insulin receptor p

ALIGNMENTS

RESULT 1

A24571

Protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human

N/Alternate names: c-erb-B-2 protein precursor; kinase-related transforming protein erbB

C/Species: Homo sapiens (man)

C/Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999

C/Accession: A24571; A25491; A44188; B44188; I59509; I57622

R/Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T.;

Nature 319, 230-234, 1986

A/Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth f

A/Reference number: A24571; MUID:86118663; PMID:3003577

A/Accession: A24571

A/Molecule type: mRNA

A/Residues: 1-1255 <YAM>

A/Cross-references: GB:X03363; NID:g31197; PIDN:CAA27060.1; PID:g31198

R/Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.

Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985

A/Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epiderm

A/Reference number: A25491; MUID:86016729; PMID:2995967

A/Accession: A25491

A/Molecule type: DNA

A/Residues: 737-1031 <SEM>

A/Cross-references: GB:M11767; NID:gi82163; PIDN:AAA35808.1; PID:g553282

R/Coussens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg, P.

Science 230, 1132-1139, 1985

A/Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chromos

A/Reference number: A44188; MUID:86070181; PMID:2999974

A/Accession: A44188

A/Molecule type: DNA

A/Residues: 740-910 <COU1>

A/Cross-references: GB:M12036; NID:gi83988; PIDN:AAA35978.1; PID:gi83989

A/Accession: B44188

A/Molecule type: mRNA

A/Residues: 1-517, 'RALL', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>

A/Cross-references: GB:M11730; NID:gi83986

R/King, C.R.; Kraus, M.H.; Aaronson, S.A.

Science 229, 974-976, 1985

A/Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.

A/Reference number: I59509; MUID:85272597; PMID:2992089

A/Accession: I59509

A/Status: translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 832-909 <REX>

A/Cross-references: GB:L29395; NID:g459807; PIDN:AAA35809.1; PID:g459808

R/Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.

Mol. Cell. Biol. 7, 2597-2601, 1987

A/Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptional

A/Reference number: I57622; MUID:87286898; PMID:3039351

A/Accession: I57622

A/Status: translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-191 <TAL>

A;Cross-references: GB:M16792; NID:g183983; PIDN:AAA58637.1; PID:g553332
C;Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
C;Genetics:
A;Gene: GDB:ERBB2; NGL; NEU; HER-2
A;Cross-references: GDB:120613; OMIM:164870
A;Map position: 17q21.1-17q21.1
A;Introns: 25/1; 75/3; 147/1; 883/3
A;Note: the list of introns is incomplete
C;Function:
A;Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
inase
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>
F;22-653/Domain: extracellular #status predicted <EXT>
F;70-304/Domain: EGF receptor extracellular domain repeat <EE1>
F;395-605/Domain: EGF receptor extracellular domain repeat <EE2>
F;654-675/Domain: transmembrane #status predicted <TM>
F;676-1255/Domain: intracellular #status predicted <INT>
F;718-983/Domain: protein kinase homology <KIN>
F;726-734/Region: protein kinase ATP-binding motif
F;168,124,187,259,530,571,629/Binding site: carboxylate (Asn) (covalent) #status predicted
F;686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F;753/Active site: Lys #status predicted
F;1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 97.4%; Score 6640; DB 1; Length 1255;
Best Local Similarity 97.5%; Pred. No. 5.5e-266;
Matches 1255; Conservative 8; Mismatches 21; Indels 2; Gaps 2;

QY 1 MELAAICRWGLLLALLPPGAASCTGCTDMKRLPASPETHDMLRLHYGQCQVQGNL 60
DB 1 MELAAICRWGLLLALLPPGAASCTGCTDMKRLPASPETHDMLRLHYGQCQVQGNL 60
QY 61 ELTYLPTNASLSFLQDIEVOGYVLIHNRQVQLRLIRVGRQLFEDNYALAVLNG 120
DB 61 ELTYLPTNASLSFLQDIEVOGYVLIHNRQVQLRLIRVGRQLFEDNYALAVLNG 120
QY 121 DPLNNTTPTVGASPGGLRELQRLSLEILKGGVLIQRNPOLCYQDTILWKDIFHKNNOLA 180
DB 121 DPLNNTTPTVGASPGGLRELQRLSLEILKGGVLIQRNPOLCYQDTILWKDIFHKNNOLA 180
QY 181 LTLIDTNRACHPCSPCKSGRWGESSEDCQSITRTVCAGGCARCKGPLPTDCCHQC 240
DB 181 LTLIDTNRACHPCSPCKSGRWGESSEDCQSITRTVCAGGCARCKGPLPTDCCHQC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNLTSTDVGSCTLVCLHNRQVTAEDGTQRCCKSPCARVCYGLGMQYIKANSKFTGIT 360
DB 301 YNLTSTDVGSCTLVCLHNRQVTAEDGTQRCCKSPCARVCYGLGMQYIKANSKFTGIT 360
QY 361 ELEFAGCKKIFGSLAFLPESPDGDPASNTAPLOPELOVFEITLGYLVIASWPDLSL 420
DB 361 IQEFAGCKKIFGSLAFLPESPDGDPASNTAPLOPELOVFEITLGYLVIASWPDLSL 420
QY 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLIGISWGLRLSRELGSGLALIHNNTHLCFVHTV 480
DB 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLIGISWGLRLSRELGSGLALIHNNTHLCFVHTV 480
QY 481 PWDQLFRPHQALLTANRPEDECVGEGGLACHQLCARGHCWGPPTQVCNCSQFLRGQEC 540
DB 481 PWDQLFRPHQALLTANRPEDECVGEGGLACHQLCARGHCWGPPTQVCNCSQFLRGQEC 540
QY 541 VEECRVLQGLPREYVNAHRLCPCHPECOPOGNSVTCFGEADQCVACAHYKDPPECVARC 600
DB 541 VEECRVLQGLPREYVNAHRLCPCHPECOPOGNSVTCFGEADQCVACAHYKDPPECVARC 600
QY 601 PSGVKPDLSPYMPIWPKPDEBACQPCPINCTHSCVDLDDKGCAPAEQASPLTISAVVG 660
DB 601 PSGVKPDLSPYMPIWPKPDEBACQPCPINCTHSCVDLDDKGCAPAEQASPLTISAVVG 660

601 PSGVKPDLSPYMPIWPKPDEBACQPCPINCTHSCVDLDDKGCAPAEQASPLTISAVVG 660
661 ILLVVLGVVFGILIKRRQKIRKYTMRLLQETVELPELTPSGAMPNQAMRILKEQYI 720
661 ILLVVLGVVFGILIKRRQKIRKYTMRLLQETVELPELTPSGAMPNQAMRILKETEL 720
721 KANSKFITITEL-TVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVGS 779
721 R-KVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVGS 779
780 PYVSRLLGICLTSTVQLVTQMPYVCLLDHVRENRGLGSQDLLNWCQIAGKMSYLEDV 839
780 PYVSRLLGICLTSTVQLVTQMPYVCLLDHVRENRGLGSQDLLNWCQIAGKMSYLEDV 839
840 RLVRDLAARNVLKSPNHVKITDFGLARLLDIDETVHADGGKVPKKNWALESLRRRF 899
840 RLVRDLAARNVLKSPNHVKITDFGLARLLDIDETVHADGGKVPKKNWALESLRRRF 899
900 THQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITDVYMWKWCW 959
900 THQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITDVYMWKWCW 959
960 MIDSECRPRELVSFMRWARPQRFVITQNEIDLGASPLDSTFYRSLLDEDDMDGLVD 1019
960 MIDSECRPRELVSFMRWARPQRFVITQNEIDLGASPLDSTFYRSLLDEDDMDGLVD 1019
1020 AEEVLVPOQGFCCPDAPAGGVMVHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSE 1079
1020 AEEVLVPOQGFCCPDAPAGGVMVHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSE 1079
1080 GAGSDVFDGDLGAAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVAPLTCSPQPEY 1139
1080 GAGSDVFDGDLGAAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVAPLTCSPQPEY 1139
1140 VNQDVRPQPPSPREGPLPAARPAATLERAKTLSPGKNGVYKDVAFAGGAVENPEYLT 1199
1140 VNQDVRPQPPSPREGPLPAARPAATLERAKTLSPGKNGVYKDVAFAGGAVENPEYLT 1199
1200 QGGAAPHPHPPAPAFDNLVYWDQDPPBERGAPPSTFKGTPTAENPEYLGLDVVP 1255
1200 QGGAAPHPHPPAPAFDNLVYWDQDPPBERGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 2
TVRTNU
Protein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 11-Jun-1999
C;Accession: A24562; A61204
R;Bargmann, C.I.; Hung, M.C.; Weinberg, R.A.
Nature 319, 226-230, 1986
A;Title: The neu oncogene encodes an epidermal growth factor receptor-related protein.
A;Reference number: A24562; MUID:86118662; PMID:3945311
A;Accession: A24562
A;Molecule type: mRNA
A;Residues: 1-1260 <BAR>
A;Cross-references: EMBL:X03362; NID:g56745; PIDN:CAA27059.1; PID:g56746
R;Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohen, C.
Carcinogenesis 12, 1975-1978, 1991
A;Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals no
2-thiazolyl]formamide or N-methyl-N-nitrosourea.
A;Reference number: A61204; MUID:92035293; PMID:1682063
A;Accession: A61204
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 637-663, 'V', 665-702 <MAS>
A;Note: authors translated the codon GCA for residue 25 as Val
C;Genetics:
A;Gene: neu
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>

F:658-680/Domain: transmembrane #status predicted <TN>
F:721-988/Domain: protein kinase homology <KIN>
F:731-739/Region: protein kinase ATP-binding motif
F:71,191,263,535,576,634/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:691/Binding site: phosphate (Thr) (covalent) #status predicted
F:758/Active site: Lys #status predicted
F:882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

Qy	1	MELAAACRWGLLLALLPGCAASTQVCTGDMKMLRLPASPTHLDMLRHLYQGQGVQGNL	60
Db	4	MELAAACRWGFLALLPGGIAGTQVCTGDMKMLRLPASPTHLDMLRHLYQGQGVQGNL	63
Qy	61	ELTYLPTNASISFLQDIQEVQGVLIHANOVQVPLQRLRIVRGTLFEDNYALAVLDNG	120
Db	64	ELTYVPANASISFLQDIQEVQGMLIHANQVKRVLQRLRIVRGTLQFEDKYALAVLDNR	123
Qy	121	DPLANNTPVT-GASFGGLRELOLSLTELKGGVLITQPNPQLCYQDTILMKDIFHKNNQL	179
Db	124	DPQDNVAASTPGRTPEGLRELOLSLTELKGGVLIRGNPQLCYQDMVLWKDVFKNQNL	193
Qy	180	ALTLLDITNRSRACHPCSPMCKSGSRGWGESSEDCQSLTRTVCAAGCARCKGFLPTDCHEQ	239
Db	184	APVDITNRSRACPPCAPACKDNHGWGESPEDCQILTGITICTSGCARCKGRLPTDCHEQ	243
Qy	240	CAAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESMPNEGRTFGASCVTAC	299
Db	244	CAAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESMHNPEGRTFGASCVTTC	303
Qy	300	PYNYLSTDVGSCITLVCPLHNQVTAEDGTQRCCKSKPCARVCYGLMGWIKANSKFIGI	359
Db	304	PYNYLSTEVGSCITLVCPPNNQVTAEDGTQRCCKSKPCARVCYGLMGHRLGARITSD	363
Qy	360	TELEPAGCKKIFGSLAFLPESPDGPASNTAPLQPEQLQVFETLEITGYLIYSAMPDSL	419
Db	364	NVQPFDGCKKIFGSLAFLPESPDGPPSSGIALPLRPEQLQVFETLEITGYLIYSAMPDSL	423
Qy	420	PDLVSFQNLQVTRGILHNGAYSLTLOGLIGISWLGRLSRELGSGLALTHNTHLCFVHT	479
Db	424	RDLVSFQNLRIIRGILHDGAYSLTQGLGIIHSLGRLSRELGSGLALTHRNAHLCFVHT	483
Qy	480	VPDQLFRNP HQALLHTANRDE-CVSGELACHOLCARGHCGWGPPTOCVNCSCQFLRGQ	538
Db	484	VPDQLFRNP HQALLHSGNRPEDLICVSSGLCVNSLCAGHCGWGPPTOCVNCSHFLRGQ	543
Qy	539	ECVEECRVLQGLPREYVNAHRLCPCHPECPQNGSVTCFGEADQCAVAHAHYKDPFCVA	598
Db	544	ECVEECRVKGLPREYVSDKRLCPCHPECPQNSSETCFGEADQCAACAHYKDSSCVA	603
Qy	599	RCPSGVKPDLSYMPWTKEPDEGACQCPDINTHSCVDLDDKGCPEAQRASPLTSTISAV	658
Db	604	RCPSGVKPDLSYMPWTKEPDESGICQPCPINTHSCVDLDERGCPAEQASPVFTFIATV	663
Qy	659	VGILLVVLGVVFGILIKRRQKIRKYTWRRLLQETELVEPLTPSGAMPNQAQMRILKEQ	718
Db	664	EGVLLFLVLVVGILLIKRRQKIRKYTWRRLLQETELVEPLTPSGAMPNQAQMRILKET	723
Qy	719	YIKANSKPIGTEL-TVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGV	777
Db	724	ELR-KVKVLGSGAGCTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGV	782
Qy	778	GSPYVSRLLGICLTSTVQLVTQIMPYGCLLDHVRENRRGLSGODLLNWCQIAKGSYLE	837
Db	783	GSPYVSRLLGICLTSTVQLVTQIMPYGCLLDHVREHRRGLSGODLLNWCQIAKGSYLE	842
Qy	838	DVRLVHRDLAARNVLKSPNHVKITDGLARLLDDTEYHADGGKVPKKNWALESLRR	897
Db	843	DVRLVHRDLAARNVLKSPNHVKITDGLARLLDDTEYHADGGKVPKKNWALESLRR	902
Qy	898	RFTHQSDVMSYCVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPPCTIDVYIMVKV	957

A:Molecule type: mRNA
A:Residues: 969-971, 'K', 973-1115, 'D' <EIS>
A:Cross-references: EMBL:Z12608
R:Heisermann, G.J.; Gill, G.N.
J. Biol. Chem. 263, 13152-13158, 1988
A:Title: Epidermal growth factor receptor threonine and serine residues phosphorylated i
A:Reference number: A28941; MUID:88330814; PMID:3138233
A:Accession: A28941
A:Molecule type: protein
A:Residues: 659-694, 'X', 696-704, 'L', 706-707;989-992, 'XX', 995-996, 'X', 998-1000;1002-1009,
R:Hibbs, M.L.; Dunn, A.R.; Alexander, W.S.
submitted to the EMBL Data Library, April 1994
A:Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor
A:Reference number: S45325
A:Accession: S45325
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-971, 'K', 973-1210 <VER>
A:Cross-references: EMBL:X78987; NID:G488830; PIDN:CAA55587.1; PID:G488831
R:Paria, B.C.; Das, S.K.; Andrews, G.K.; Dey, S.K.
Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993
A:Title: Expression of the epidermal growth factor receptor gene is regulated in mouse b
A:Reference number: I49643; MUID:93126380; PMID:7678348
A:Accession: I49643
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 12-20, 22-132 <RES>
A:Cross-references: GB:L06864; NID:G193001; PIDN:AAA53029.1; PID:G567201
C:Genetics:
A:Gene: EGFR
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphop
F:1-24/Domain: signal sequence #status predicted <SIG>
F:648-670/Domain: transmembrane #status predicted <TM>
F:712-977/Domain: protein kinase homology <KIN>
F:720-978/Region: protein kinase ATP-binding motif
F:680,695/Binding site: phosphate (Thr) (covalent) #status experimental
F:697,1070,1071/Binding site: phosphate (Ser) (covalent) #status experimental
F:993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental
F:1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental
F:1197/Binding site: phosphate (Tyr) (covalent) #status experimental

Query Match	44.6%; Score 3038; DB 2; Length 1210;
Best Local Similarity	49.1%; Pred. No. 6.7e-118;
Matches	625; Conservative 167; Mismatches 368; Indels 114; Gaps 27;
Qy	11 LLLALLPPGAA--STQVCTGTDMKLRUPASPETHLDMLRLHYQCCVQVQGNLELYLPTN 68
Db	14 LLTRLCAGGNALEKKVCOQTSNRLTQLTGFEDHFLSLQRYNNCEVVLGNLEITYYQVN 73
Qy	69 ASLSFLQDIQEVQGYVLIAHNQVQVPLQRLRIVRGTQLFEDNYALAVLQNGDPLNNTTP 128
Db	74 YDLSFLKTIQEVAGYVLIALTNTVERIPLENLQIIRGNALYENTYVALILSN----- 124
Qy	129 VTGASPGCLRELQRLRSITELKGGVLTQRPOLCYQDTILWKDI-----PHKNQALATLI 184
Db	125 -YGTNRITGLRELPMRNIQETLIGAVRFSPNNPILCNMDTIQWRDTIQVQVPSNMSMDL--- 180
Qy	185 DTRNSRACHPCSPMCKSGSRGMSSEDCQSLTRTWCAGGCA-RCKGPLPTDCCHCEQCAAG 243
Db	181 -QSHPSSCPKDCSPNGSCWGGEENCQKLTKLIQAQCSHRRCGRSPSDCCHNQCAAG 239
Qy	244 CTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFBSMPNPEGRYTFGASCVTACPNY 303
Db	240 CTGPRESDECLVQCFQDEATCKDTCPPMLMLNPTTYQMDVNPEKYSFGATCVKKCPRY 299
Qy	304 LSTDVGSCCTLVCLPHNOVTAEDGTQCEKSKPCARVCYGLGMQYIKANSKFITIGITELE 363
Db	300 VTDHSGSVACGPDYIEV--EEDGIRCKKCDGFCRVCNGIGIGEPK-DTSLINATNIK 357
Qy	364 -FAGCKKIFGSLAPSPESFDGDPASNTAPLQPEQLQVFTLEETIGLYLISAWPDSLPDL 422
Db	358 AHFYCTAISGLDHLIPVAFKGSFSTRPDLQPRELEILKTKVETLQGLLQAWPDNDTDL 417

Qy 423 SVFQNLQVIRGRILHNGAVSLTLOGLGISWLGRLSRLBELSGSLALHNNTHLCFVHTVPV 482
 Db 418 HAFENLEIIRGRTKQHGQSFSLAVVGLNITSLGRSLIKEISDGVIIISGNRNLCYANTINW 477
 Qy 483 DQLFRPHQALLHTANRPEDECVGEGLACHOLCARGHCWGPPTQCVCNSQSFURGOBCEV 542
 Db 478 KKLFGTPTNOKTKIMNNAEKDCKANVHVCNPLCSSEGCWGPPEPDCVSCQNVSGREGCEV 537
 Qy 543 ECRVLOGLPREYYNARHCLPCHPECOFQNGSVTCFPEADQCACAHYKDPFPCVACRPS 602
 Db 538 KCNILEGEPREFVENSECICQHEPECLPQAMNITCTGRGPDNCICQAHYIDGPHCVKTCPA 597
 Qy 603 GVXPDLISYMIWKPFPDEBAGACQCPNCTHSCVDLDDKGPAPQORASPLTSTISVAVVGIL 662
 Db 598 GINGENNTL-VWKYADANNVCHLCHANCTYGCAGPLQGCCEVMPSPGKIPSIATGIVGGL 656
 Qy 663 LVVVLGVVFGI-LIKERQOKIRKYWMRRLLQELVEPLTPSGAMPNQAOAMRLIKEQYIK 721
 Db 657 LFIVV-VALGIGLFWMRHHIVRKTIRRLLORELVEPLTPSGEAPNQAHRLIKETFEK 715
 Qy 722 ANSKFTIGITEL-TVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVSP 780
 Db 716 -KIKVLGSGAGFTVYKGLMPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDP 774
 Qy 781 YVSRLLIGICLTSTVOLVTQMLMPYGCLLDHVRENRGRGLSQDLLNWCQIAGKHSYLEDDR 840
 Db 775 HVCRLLGICLTSTVOLITQMLMPYGCLLDVYREHKDNIGSQYLLNWCQIAGKNWYLEDRR 834
 Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETETHADGGKVPKMWMALESILRRRFT 900
 Db 835 LVHRDLAARNVLKTPQHVKITDFGLAKLGAEEKEYHAEKKVPKMWMALESILHRIYT 894
 Qy 901 HQSDVMSYGVTVWELMTFGAKYDGPAREIPDLLEKGERLPQPPICITDVYIMVVKCWM 960
 Db 895 HQSDVMSYGVTVWELMTFGSKPYDGPASDISILEKGERLPQPPICITDVYIMVVKCWM 954
 Qy 961 IDSECPRRPRELVSESRMARDQRFVITQ-NEDLGPASPLDSTFYSRLLEDDEDDMGDLVD 1019
 Db 955 IDASRPPKRELILEFSEQWARDQRYLVITQGBRMHLPSPTDSNFYRALMEDMEDMEDVD 1014
 Qy 1020 AEYLYVPOQGFPCPDAPGAGGVHHRHSSYTRSGGDLTLGLEPSEEAAPSPAPSE 1079
 Db 1015 ADYLYLPQGGFF-----NSPST-----SRTPLLSL 1040
 Qy 1080 GAGSDVFDGLGMAAKGLOSLFTHDPSPLQRYSEDDPTVPLPSET--DGYYAPLTCSPQP 1137
 Db 1041 SATSN---NSTVACINRRSGCRVKEDAFQORYSSDPTGAVTEDNDDAPL-----PVP 1090
 Qy 1138 EYVNOQDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKGVVVDVAFPGGAVENPEYL 1197
 Db 1091 EYVNO-SVPKRPAGSVQNPVYHNOPLHP-----APGRDLHYQN--PHSNVAVGNPEYL 1139
 Qy 1198 -TPQGGAAQPPHPHPAPFAPFADNLYWQO-----DP-----PERGAPPSFPGT 1240
 Db 1140 NTAQ-----PTCLSSGFNSPALWIKGSHQMSLDNPDYQDDFPFKETKNGLFGK- 1189
 Qy 1241 PTAENPEYLGLDVP 1254
 Db 1190 PTAENAEYLRVAPP 1203

RESULT 6
 TVCHLV
 epidermal growth factor receptor precursor - chicken
 N:Contains: protein-tyrosine kinase (EC 2.7.1.112) etbb
 C:Species: Gallus gallus (chicken)
 C:Date: 28-Feb-1986 #sequence_revision 05-May-1995 #text_change 04-Feb-2000
 C:Accession: A27720; A00643
 R:Lax, I.; Johnson, A.; Howk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.
 Mol. Cell. Biol. 8, 1970-1978, 1988
 A:Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression
 A:Reference number: A27720; MUID:88261272; PMID:3360329

A:Accession: A27720
 A:Molecule type: mRNA
 A:Residues: 1-1223 <LAX>
 A:Cross-references: GB:M20386
 R:Nilsson, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
 Cell 41, 719-726, 1985
 A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
 A:Reference number: A00643; MUID:85228222; PMID:2988784
 A:Accession: A00643
 A:Molecule type: mRNA
 A:Residues: 585-1223 <NIL>
 A:Cross-references: GB:M10066
 C:Genetics:
 A:Gene: erbB
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
 specific protein kinase
 F:1-30/Domain: signal sequence #status predicted <SIG>
 F:31-1223/Product: epidermal growth factor receptor #status predicted <MAT>
 F:31-654/Domain: extracellular #status predicted <EXT>
 F:81-307/Domain: EGF receptor extracellular domain repeat <EE1>
 F:397-610/Domain: EGF receptor extracellular domain repeat <EE2>
 F:655-677/Domain: transmembrane #status predicted <TM>
 F:678-1223/Domain: intracellular #status predicted <INT>
 F:719-984/Domain: protein kinase homology <KIN>
 F:727-735/Region: protein kinase ATP-binding motif
 F:136,202,280,361,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent) #
 F:192,650/Binding site: carbohydrate (Ser) (covalent) #status predicted
 F:687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
 F:754/Active site: Lys #status predicted
 F:1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat

Query Match 44.2%; Score 3014.5; Db 1; Length 1223;
 Best Local Similarity 47.9%; Pred. No. 6.2e-117;
 Matches 622; Conservative 173; Mismatches 355; Indels 149; Gaps 29;

Qy 8 RWGILLALLPGAA-----STQVCTGTMKRLRIPASPETHLDMLRHLYQCQVVOGNLE 61
 Db 13 RGAALVLLGLGVALCSAVEKKVCQGTNNKLTQLGHVEDHFTSLQRMNNCEVVLNLE 72
 Qy 62 LTYLPTNASTLFDIQLQVGVVLIANNQVQVPLQRLRIVRGTQLPEDNYALAVLNDG 121
 Db 73 ITYEHNRLDTFLKTIQVAGVGLIANNMVDVPLENQLIRGNVLYDNSFALAVLSNH 132
 Qy 122 PLNNTTPVTGASPGGLRLQLRSLEILKGGVLIQNPOLCYQDTILWKDIFHKNOLAL 181
 Db 133 -MNKTO-----GLRLPMKRLSEILNGVVKISNNPKLNMVTLWNDIIDTSRK-PL 182
 Qy 182 TLID-TNRSRACHPCSPMKSGRCWGESSEDCQLTRTVCAAGCA-RCKGPLPTDCCHEQ 239
 Db 183 TVLDFASNLSSCPKCHPNCTEDHCWAGEQNCQTLTKVICAQCSGRCGRKVPSDCCHNQ 242
 Qy 240 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVYNTDTFESMPNPEGRTYFGASCVTAC 299
 Db 243 CAAGCTGPGESDCLACRFDRDATCKDTPPLVLYNPTTYQMDVNPBGKYSFGATCVREC 302
 Qy 300 PYNLYSTDVGSGTLVPLPHNEVTAEDCTQCEKSKPCARVCYGLGMOYIKANSKFTGI 359
 Db 303 PHNYVTDHSGVRSVCNTDTEV-EENGVRCKCKDGLCSKVCNGIGELKGLIS-INA 360
 Qy 360 TELE-FACGKKIFGSLAPSPDGPASNTAPLQPEQLQVFTLEITGYLYISAWPDS 418
 Db 361 TNIDSFKNCTKINGDVSILPVAFLGDAFTKTLPLDPKLDVFTVKEISGFLLIQAWPDN 420
 Qy 419 LPDLVSFONLQVIRGRIILHNAGYSITLQGLIGISWGLRSRELGLALIHNTLHLCFVH 478
 Db 421 ATDLYAFENLEIIRGRTRKHQGYSLAVNVLKIQSLGLRSLKEISDGDIAIMKNKLCYAD 480
 Qy 479 TVPMDQLFRNPQALLHTANPEBCEGCEGLACHQLCARGHCWGPPTQCVCNCSQFLRGQ 538
 Db 481 TNWRSFLPATOSQTKTIQNRKNKNDCTADRHVDPLCDVCGWGGPGPHCFSCFFSRQK 540
 Qy 539 ECVBEVRVLOGLPREYVNAHCLFCHPECPQNG---SVTCFGEADQCVACAHYKDPFF 595

Db 541 ECVKQCNILQGEPRFERDSKCLPCHSECLVQNSTAYNTTCSGPGPDHCKMKCAFIDGPH 600
 Qy 596 CVARCPGVPDLISYMPIWKFPPDEEGACQPCPINCTHSCVDLDDKQGAEPAGRAQSPITSIV 655
 Db 601 CVKACPAVVGENDTL-VWKYADANAVCQLCHPNCTRGCKGPGLEGCP---NGSKTPSIA 656
 Qy 656 SAVV-GILLVVVLGWVFGIILKRRQKIRKYTWRLRLQETELVEPLTFPSGAMPNQAOMRI 714
 Db 657 AGVVGLLCLVWGLGIGLYLRRR-HIVRKRTLRLLAQERELVEPLTFPSGEAPNQAHLRI 715
 Qy 715 LKEQYIKANSKFTIGTEL-TYKGIWIPDGENVKIPVAIKVLRENTSPKANKELIDRAYV 773
 Db 716 LKETEPR-KYKVLGSGAFGVYKGLWPEGEKIPVAIKELREATSPKANKELIDRAYV 774
 Qy 774 MAGVSPVVRLLGICLTSTVQLVTLQMPYGCLLDHVRENKRLGSDGLLNCWQCIKAGM 833
 Db 775 MASVDNPHVCRLLGICLTSTVQLITQUMPYGCLLDYIREHKDNIGSYLLNWCVOIKAGM 834
 Qy 834 SYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIKWMALES 893
 Db 835 NYLEERLVRDLAARNVLKTPQHVKITDFGLAKLIGADEKEYHAGSGKVPKIKWMALES 894
 Qy 894 ILRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPTAREIPDLLEKGERLPPOPICTIDVYM 953
 Db 895 ILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPTASEISSVLEKGERLPPOPICTIDVYM 954
 Qy 954 IMVKCMWIDSECPREELVSESRMARDPQRFVVIQ-NEDLGPASPLDSTFVRSLLAEDD 1012
 Db 955 IMVKCMWIDADSPPKRELIABFSKWARDPPRYLVIOGQDERMHLPSPTDSKFYRTLMEE 1014
 Qy 1013 DMGLVDABEYLVPOQFFCPDPAPGAGVMVHRHRSSTSTRSGGDLTLGLEPSEEEAPR 1072
 Db 1015 DMEDIVDADEYLVPHQGF-----NSPST-----SR 1040
 Qy 1073 SPL-----AFSEGAGSDVFDGLGMGAAGLQLSPLTHDPSPQLQRYSDPTVPLPSET-D 1125
 Db 1041 TPLLSSLATSNNASATNCID-----RNGQGHVREDSFVQRYSSDPTGNFLEESIDD 1092
 Qy 1126 GYVAPLTCSPQEVVNPQDVRPQPPSPREGPLPAARAGATLERAKTSLSPKGVGVKDV 1185
 Db 1093 GFL-----PAPEYVNO--LMPKKPS-----TAMVQNIYNNIS 1123
 Qy 1186 -----AFGGAIVENPEYLTPOGGAAPPHPPPAFSPAFDNLVYMDQ----- 1225
 Db 1124 LTAISKLPMDSRVQNSHSTAVDNPYL-----NTNQSFLLAKTVPESSPYLQSGNHQ 1175
 Qy 1226 ---DPPB-----RGAPPSTFKGTPTAENPEYLGLDVP 1254
 Db 1176 INLDNPDYQQLFPLNETKPNKNGLLKVPAAENPEYLRVAAP 1214
 RESULT 7
 A:47253
 C:epidermal growth factor receptor, HER4 - human
 C:Species: Homo sapiens (man)
 C:Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
 C:Accession: A47253
 R:Plowman, G.D.; Culouscou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.; Ne
 Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993
 A:Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epidermal
 A:Reference number: A47253; MUID:93189574; PMID:838336
 A:Accession: A47253
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: nucleic acid
 A:Residues: 1-1308 <PLO>
 A:Cross-references: GB:L07868; NID:g337359; PIDN:AAB59446.1; PID:g337360
 A:Note: sequence extracted from NCBI backbone (NCBIP:126842)
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; growth factor receptor
 F:716-981/Domain: protein kinase homology <KIN>
 F:724-732/Region: protein kinase ATP-binding motif

Db 123 YQK-NPSSP--DVYQVGLKQLQLNLNLTEILSGGVKVSHPNLLCNVETINWWWDIVDKTSNP 179
Qy 180 ALTLLIDNRSRACHPCSMCKGSRGWESSBDCSLTRTVCAAGC-ARCKGPLETDCCHE 238
Db 180 TNNLI PHAFERQCKDHCVCNGSWAPGPGHCQKFTKLLCAEQCNRRRCRGPFPIDCNE 239
Qy 239 QCAAGCTGPKSHDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRTTFGASCUTA 298
Db 240 HCAGGCTGPRATDCLACRDFNDGCTCKTCTPPPKIYDIIVSHVQVVDNPNIKYTFGAACVKE 299
Qy 299 CPYNVLSTDVGSCTLVCPHLHNOEYTAEDGTORCEKSKPCARVCYGLGMQVIKANSKFIG 358
Db 300 CPSNVVTE-GACVRSACSAGMLEVD-ENGKRSKPCDGVCKVCDGIGISL-SNTIAVN 356
Qy 359 ITEL-EFAGCKKIFGSLAFLESFSDGDPASNTAPLOPEQLQVFTLEEBITGYLISAWPD 417
Db 357 STNIRSFNSCKINGDIILNRNSFEGDPHYKIGTMDPEHLWNLTIVKEITGYLVIMWPE 416
Qy 418 SLPLSVFQNLQVTRGRILHNGAYS-LTLOGLGISWLGRLSLRELGLSLALIHNTLHCF 476
Db 417 NMTSLSVFQNLLEIIRGRTTFRGFSFVVVQVRHLQWLGLRLSKVSAGNVILKNTLQRY 476
Qy 477 VHTVPWDLRPNHQAALLHTANRPEDECVGEGLAHQLCARGHCWGPCTOCVNCQFLR 536
Db 477 ANTINWRFLRSEDSQSEYDART-----ENQTCNNECSDGCGW-FGPTMVCVSLHVD 528
Qy 537 GQECVECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFPGPADOCVCAHAHYKDPFFC 596
Db 529 GGRVCASCNLLQGEPRQAQVDRGCVOHQECLVQTDLSLTCVPGPANCKSAHFQDGPQC 588
Qy 597 VARPSPGVKPLSYMPYKFWPDEBAGCQPCPINCTHSCVDLDDKGCBAQASPLTIVS 656
Db 589 IPRCPHGILGDGTL-IWKYADKMGCQCPCHQNCQTQCGSGPLSGCRGD-IVSHSLAVG 646
Qy 657 AVVGLLVVVLGVVFGILIKEROOKIRKYWRRLLOETELVEPLTPSGAMPNQOMILK 716
Db 647 LVSGLLTIVALLIVLILRRRRIRK-RKRTIRCLLQEKELVEPLTPSGAQNPAFLRLK 705
Qy 717 EOYIKANSKFIGITEL-TVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYMA 775
Db 706 ETEFK-KDRVLGSGAFGVYKGLWNPDGENIRIPVAIKVIREATSPKVNQEVLEAYMA 764
Qy 776 VGSPFYSRLIGICLTSTVQLVTOLMPYGCGLLDHVRENRLGSDQLLNWCQIAKMSY 835
Db 765 SVDPHPVCRLLIGICLTSAVQLVTOLMPYGCGLLDVYRQHOERICGOWLLNWCVOIAKMN 824
Qy 836 LEDVRLVHRDLAARNVLKSPNHVKITDPGLARLLDDETHYADGGKVPKKNWALSIL 895
Db 825 LEERHLVHRDLAARNVLKSPNHVKITDPGLARLLDDETHYADGGKVPKKNWALSIL 884
Qy 896 RRRPTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIM 955
Db 885 QWYTHOSDVMSYGVTVWELMTFGSKPYDGIIPAREIASVLENGERLPPOPICTIEVMI 944
Qy 956 VKCWMIDSECRPRRELVSFERSWARDPQRFVWQNEEDLGPASPLDSTFYRSLLEDMDMG 1015
Db 945 LKCMWIDPSSRPRRELVGESQWARDPSRYLVQ---NLPSSLDRFLRSLSSDD-- 999
Qy 1016 DLVDAEYLVYVQOQFFCPDAPAGAGGVHHRHSSSTRSGGDDTLGLEPSEEAAPRSL 1075
Db 1000 DVVDAEYLVYVYKRI-----NRQGS-----EPC 1022
Qy 1076 APSEGAGSDVPDGLGMAAGLQSLPHTDPSLPQRYSEDPV-PLPSETDGYVAPLTC 1134
Db 1023 IPPTGH-----PVRENSITLRNISDPTQNALEKLDLGH----- 1055
Qy 1135 POPEYVNPQDVRPQ-----PSPRE-----GPLP-AARPAGATLERAKTLPFGKNGV 1181
Db 1056 ---EYVNPQSGSETSSRLSDIYNPNVEDLTDGWPVPSLSSQEAETNFSRPEYLNQNSL- 1111
Qy 1182 KDVFAGGAVENPEYLPQGAAPQPHPPAFSPAFNLYWDOQDPPERGAAPSTFKGTP 1241
Db 1112 --PLVSSGMDPDY---QAG-----YQAAF-----LPQTGALTGNGMFLP 1147

Qy 1242 TAENPEYLK 1250
Db 1148 AAENLEYLG 1156

RESULT 9

A36223
C:Species: Homo sapiens (man)
C>Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000
C/Accession: A36223; I59164
R/Kraus, W.H.; Issing, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.
Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989
A>Title: Isolation and characterization of ERBB3, a third member of the ERBB/epidermal g
A/Reference number: A36223; MUID:90083234; PMID:2687875
A/Accession: A36223
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-1342 <KRA>
R/Plowman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todaro, G.J
Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990
A>Title: Molecular cloning and expression of another epidermal growth factor receptor-re
A/Reference number: I59164; MUID:90311312; PMID:2164210
A/Accession: I59164
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-559, 'G', 561-957, 'F', 959-1063, 'G', 1065-1342 <RES>
A/Cross-references: GB:M34309; NID:G183990; PID:AAA35979.1; PID:G306841
C/Genetics:
A/Gene: GDB:ERBB3; HER3
A/Cross-references: GDB:119880; OMIM:190151
A/Map position: 12q13-12q13
C/Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog
C/Keywords: ATP; phosphotransferase
F/707-972/Domain: protein kinase homology <KIN>
F/715-723/Region: protein kinase ATP-binding motif

Query Match 34.2%; Score 2332.5; DB 2; Length 1342;
Best Local Similarity 39.5%; Pred. No. 7.5e-99;
Matches 520; Conservative 197; Mismatches 461; Indels 137; Gaps 36;

Qy 10 GLLLALLPPGAA--STQVCTGTDMKRLRSPETHLDMRLHLYQGCQVQGNLELYLPT 67
Db 11 GLLFSLARGSEVGNQAVCPGTLNGLSVTGDENQYQTLKYERCEVVMGNLEIVLTGH 70
Qy 68 NASLSFLQDIQEVQGVVLIHNNQVRQVPLQRLRIVRGTLQFEDNYALAVLDNGDPLNNT 127
Db 71 NADLSFLQWTFEVTGYVYLVAMNEFSTLPLNLRVVRGTQVYDGKFAIFVM-----LYNNT 125
Qy 128 PVTGASPGGLRELQARSLTEILKGGVLIQNPOLCYODTILWKDIFHKNNQLALTLDTN 187
Db 126 ----NSSHAURQLRLTQLTEILSGGVYIEKNDKLCHEMDITDWRD---AEIVVKD 178
Qy 188 RSRACHPCSPMCKGSRGWESSBDCSLTRTVCAAGC-ARCKGPLETDCCHECAAGCTG 246
Db 179 NGRSCPPCHEVCKG-RCWGPSEDCQTLTKTICAPQCNHGCFGNPNQCCHDECAGCGS 237
Qy 247 PKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRTTFGASCVTACPYNYLST 306
Db 238 PQDTCDFACRHFNDGACVPRCPQPLVYNKLTQLEPNPHTKYQYGGVCVASCAPHNEV-V 296
Qy 307 DVGSCSTLVCPHLHNOEYTAEDGTORCEKSKPCARVCYGLGMQVIKANSKP--IGITELE- 363
Db 297 DQTSQVRCAPPDQKMEVD-KNGLKWCPCPGGLCPKACEGTG-----SGSRFQTVDSNNID 350
Qy 364 FAGCKKIFGSLAFLESFSDGDPASNTAPLOPEQLQVFTLEEBITGYLISAWPDSLPLDS 423
Db 351 FVNCTKILGNLDELITGLNGDPWHKIPALDPEKLVNFRVTREITGYLNTQSWPPHMNF 410
Qy 424 VFQNLQVIRGRILHNGAYS-LTLOGLGISWLGRLSLRELGLSLALIHNTLHCFVHTVPW 482

Db 411 VFSNLTITIGRSLYNRGFSLLIMKNLNTSLGFRSLKEISAGRIYISANRQLCYHHSLSNW 470
Qy 483 DOLFERNHQAALLHTA-NRPEDECVGEGGLACHQLCARGHCWGPPTQCVCNCSQFLRGQECV 541
Db 471 TKVLRGPTFEERLDIKHNPRDCAVAGKVDPLCSSGCGWGPQCCLSCRNRYGRGVGV 530
Qy 542 EECRVLOGLPREYVNAHCLCHPECOQPONGSVTCFPGPEADQCACAHYKDPFPCVARCP 601
Db 531 THCNFLNGEPREFAEAEFCFCHPECOQMEGTATCNGSGSDTCAQCAHFRDGHCVSSCP 590
Qy 602 SGVXPDLUSYMPIWKPFPDEGACQPCINCHTSCVDLDDKGPAPBORA-----SPLTISUSA 657
Db 591 HGVLG--AKGPIKYKYPQVQNECRPCHENCHCTQCGKPELQDCLGTGLVLIKTHLTALTY 648
Qy 658 VVGILLVVVGVLGILIKRQOKIR-KYTWRRLLQETELVEPLTPSGAMPNQOMRLK 716
Db 649 IAG--LVVIFWMLGGTFLYWRGRRIONKRAMRRYLERGESIEPLDPS-EKANKVLARIFK 705
Qy 717 EYIKANSKFTIGTEL-TVYKGIWIPDGENVKIPVAIKVLRNTSPKANKILDEAYMA 775
Db 706 ETELK-KLVKLGSGVGTGKGVWIPGESIKIPVCIKVIEDKSGRSQFQAVTDHMLAIG 764
Qy 776 GVGSPYVRLGICLTSTVQLVTPQMPYGCILLDHNVRNRLGSDLLNWCQIAKMSY 835
Db 765 SLDHAHIVRLGLCPGSSQLVTVPLGSLLDHVRQHRGALGPQLLLNMGVQIAKMWY 824
Qy 836 LEDVRLVHRDLAARNVLKSPNHVKITDPGLARLLDDETEYHADGGKVPKIKWALSIL 895
Db 825 LEEHGMVHRNLAARNVLKSPQVADFVADLLPPDDKQLLYSEAKTPIKMWALSIEH 884
Qy 896 RRRFTHQSDVMSGYVTVMELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPTCTTIDVYIM 955
Db 885 FGKTHQSDVMSGYVTVMELMTFGAEPYAGLAEVDPDLLEKGERLAQPOICTIDVYVM 944
Qy 956 VKCMWIDSECRPRFELVBSFMRWARDQRFVVIQNEDLGPA---SPLDSTFYSLLEDD 1012
Db 945 VKCMWIDENIRTEKELANEFTRMARDPPRYLVIKRES-GPGIAPGPPHGLTKKLEEV 1003
Qy 1013 DMGDLVDAEEVLVPOQFFCDPAPAGCMVHHRSSSTRSGGDLTLGLEP-SBEEAP 1071
Db 1004 ELEPELDLDLEAED-----NLATTLGSAUSLPGTLNLRPGS 1044
Qy 1072 RSLAPSGAGSDVDFDGLMGAAGLQSLPTH-D-PSPLQRYSEDPVTPLP-----SET 1124
Db 1045 QSLSPSGY-MPMNQNLGSCQESAVSGSERCPRVSLH-----PMRPGCLASES 1097
Qy 1125 DGYVA-----PLTCSPOPE-----YVNPQDVVRPPOPPSPREGP----- 1156
Db 1098 EGHVTSBAELQEKVSMCRSRSRSPRGRDSAYHSQRHSLLTPVTPLSPGLLEEDVN 1157
Qy 1157 ---LPAARPAGATLERAKTISP-GKNGV-----KDVFAEGGAVENPEYLTPOGGAPO 1206
Db 1158 GYVMPDTHLKTGPSREGTSLSVGLSSVLGTEEDED-----EYENYNNRRRHSR- 1208
Qy 1207 PHPPAFAPFADNLVYMD-----QDPERGAPPSTFKGTPTAENPEYL 1249
Db 1209 PHPPRPSLEBELGYEYMDVGDLSASLGSTQSCPLHPVIMPTAGTITPDEDIETM 1263

RESULT 10
JC4387
epidermal growth factor receptor homolog precursor - rat
N:Alternate names: ErbB3 protein; HER3 protein
C:Species: Rattus norvegicus (Norway rat)
C:Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998
C:Accession: JC4387
R:Hellyer, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.
Gene 165, 279-284, 1995
A:Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protein.
A:Reference number: JC4387; MUID:96096535; PMID:8522190
A:Accession: JC4387
A:Molecule type: mRNA
A:Residues: 1-7339 <HEL>

A;Cross-references: GB:U29339; NID:G915389; PID:G915390
A;Experimental source: liver
A;Note: The authors translated the codon AAC for residue 369 as Thr and GTT for residue 370
C;Comment: This protein is a functional heregulin receptor that transduces signals to the
C;Genetics:
A;Gene: ErbB3
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog
C;Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane protein
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1339/Product: epidermal growth factor homolog #status predicted <MAT>
F:640-659/Domain: transmembrane #status predicted <TM>
F:705-970/Domain: protein kinase homolog <KIN>
F:713-721/Region: protein kinase ATP-binding motif
F:939,1051,1156,1194,1196,1219,1257,1259,1273,1286,1325/Binding site: phosphate (Tyr) (cc
Query Match 33.0%; Score 2246.5; DB 2; Length 1339;
Best Local Similarity 39.7%; Pred. No. 2.6e-85;
Matches 511; Conservative 173; Mismatches 441; Indels 161; Gaps 37;
Qy 3 LAALCRWGLLLALLPPGAA---STQVCTGTDMLRLPASPTHLDMLRLHLYQCGQVVOGN 59
Db 7 LQVLC-----FLLSLARGSEMGNSQAVCPGTNLGSLVTGDDADNQYQTLKLYKECEVWGN 62
Qy 60 LETYLTPTNASLSFLQDIOEVQGYVLIHNRQVPLQRLRIYVRGTQOLFEDNVALAYLDN 119
Db 63 LEIVLTGHNADLSFLQWIREVTAYVLVAMNEFSVLPLENLRVVRGTQVYDQKFAIFVM-- 120
Qy 120 GDPINNTPTVGASPGRLRLQLRLTEILKGVLTORNPOLCYQDTILWKDIFHKNNQL 179
Db 121 ---LNYNT-----NSHALRQLKFTQLTEILSGGVYIEKNDKLCMDTIDRIVRVR--- 170
Qy 180 ALTLIDNRSRACHPCSPCKSGRSGWESSEDCQSITRTVCAGC-ARCKGPIPTDCHE 238
Db 171 GASIVKNGANGCPCHEVCCKG-RCMGPGDDCQILITIKAPQCNCRGCFGNPQCCHD 229
Qy 239 QCAAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDITFESMPNPEGRYTFGASCVTA 298
Db 230 ECAGGCGPQDTCFACRRFENDSGACVPRCPPELVYNTKLTFOLEPNHTKYQYGGVCVAS 289
Qy 299 CPYNYLSTDVGSCTLVCPHLNQVETADGTORCEKSKPCARVCYGL--GMQYIKANSKF 356
Db 290 CPNHFV-VDQTFVCRACPPDDMEVD-KHGLKMCPCGGLCPKACEGTSGSRYQTVDSN 347
Qy 357 IGITELEFACCKIFGSLAFLPESFDPDASNTAPIQEOLOVFELEETITGLYLTISAWP 416
Db 348 ID---GFVNCTKILGNLDFLITGLNVDPWHKIPALDPEKLVNFRVREITGLYNTQSWP 403
Qy 417 DSLPDLISVFONQIVIRGRIHNGAYS-LTLQGLGISWGLRSLRSLRSLGSLALIHNTSLC 475
Db 404 PMHNFVSFNLTTIGRSLYNRGFSLLIMKNLNTSLGFRSLKEISAGRIYISANQOLC 463
Qy 476 FVITVPWDQLFRPHQALLHTA-NRPEDECVGEGGLACHQLCARGHCWGPPTQCVCNCSQ 534
Db 464 YHSLNWTLLRGPSEERLDIKYDRPLGELAEKGVCDPLCSCSGCGWGPAGQCLSCRNY 523
Qy 535 LRQECVEECRVLOGLPREYVNAHCLCHPECOQPONGSVTCFPGPEADQCACAHYKDP 594
Db 524 SRGVCVTHCNFLQGGPREFEVHEAQCFSCHEPCLPMEGTSTYNGSGSDACARCAHFRDGP 583
Qy 595 FCVARCPGKVPDLISYMPIWKFPDEGACQPCINCHTSC--VDLDDKGPAPBQASPLT 652
Db 584 HCYNSCPHTGLG--AKGPIKYKYPDAQNECHPCENCTQGCNGPELQDCLQAEVLSMKPH 641
Qy 653 SIYSVAVVGIILLVVVLGVVFGILIKRQOKIR-KYTWRRLLQETELVEPLTPSGAMPNQAO 711
Db 642 LVIAVTVG--LAVILMLGGSFLYWRGRRIONKRAMRRYLERGESIEPLDPS-EKANKVL 698
Qy 712 MRILKEQYIKANSKFTIGTEL-TVYKGIWIPDGENVKIPVAIKVLRNTSPKANKILDE 770
Db 699 ARIFKETELR-KLVKLGSGVGTGKGVWIPGESIKIPVCIKVIEDKSGRSQFQAVTDH 757
Qy 771 AYVMAGVSPYVSRLLGICLTSTVQLVTPQMPYGCILLDHNVRNRLGSDLLNWCQIA 830

Query Match 23.8%; Score 1623; DB 1; Length 604;
Best Local Similarity 50.7%; Pred. No. 5.1e-60;
Matches 350; Conservative 76; Mismatches 137; Indels 128; Gaps 18;

QY 587 CAHYKDPFCVACPSGVKPDLSYMPIWKFPDEBAGACQPCPINCTHSCVDLDDKGCAPAEQ 646
DB 3 CAHFTDGHCHVKACPAVGLGENDTL-VKRYADANAVCQLCHPNCNCTRGKPGLEGCP--- 58

QY 647 RASPLTSIVSAV-GILLVVVLGVVFGILLIKRQOKIRKYIMRRLLQETELVEPLTPSGA 705
DB 59 NGSKTPSTAAGVVGGLCLLVVGLGIGLYLRR-HIVAKRTLRRLQORELVEPLTPSGE 117

QY 706 MPNOAMRILKEQYIKANSKFTIGTEL-TVYKGIWIPDGENVKIPVAIKVLRNTPSKAN 764
DB 118 APNOAHLILKETEFK-KVKVLGSAFGTYKGLWIPSEKVKIPVAIKELREATSPKAN 176

QY 765 KEILDEAYMAGVGSYVSRLLGICLTSTVOLTMYPYGCGLLDHVRNRRGLSQDILLN 824
DB 177 KEILDEAYMASVDNPHVCRLLGICLTSTVOLITQMPYGCGLLDYIREHKDNIGSQYLLN 236

QY 825 WCMQIAKMSYLEVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEYHADGGKV 884
DB 237 WCVQIAKGMNYLEERLVRDLAARNVLKTPQHVKITDFGLAKLLGGADEKEYHAEGKV 296

QY 885 PIKMALESILRRRFTTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQ 944
DB 297 PIKMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIPIASEISSVLEKGERLPQ 356

QY 945 PICTIDVYIMVWKMIIDSECRPRELVSEFSRMARDPQRFVVIQ-NEDLGPASPLDST 1003
DB 357 PICTIDVYIMVWKMIIDADSRPKFRELIAEFKSMARDPPRYLVITQDERMHLPSPTDSK 416

QY 1004 FYRSLLEDDMDLVDAEYLVPOGFFCPDPAFCAGGQVWHRHRSSTRSGGGLDITGL 1063
DB 417 FYRTLMEBEDMEDIVDAEYLVPHQGF-----NSPST----- 449

QY 1064 EPSEEEAPRSP-----APSEGAGSDVDFDGLGMAAGLQSLPHTHDPSPQRYSEDPV 1118
DB 450 -----SRTPLLSLSATSNNSATNCID-----RNGQHPVREDSFQRYSSDPTG 494

QY 1119 PLPST-DGVYAPITCSPQBYVNPQVDRPQPPSPREGPLPAARPAAGATLERAKTLSPG 1176
DB 495 NPLESIDDGLF-----PAPEYVQ--LMPKKPSTAM----- 524

QY 1177 KNGVVKVFAF-----GGAVENPEYLPQCGAAPQHPHPPAFSPAF 1217
DB 525 ---VQNYINFLISLTAISKLPMSRYQNSHSTAVDNPEYL-----NTQSPILAKTVF 573

QY 1218 DNLVYWDQPPPERGAPPSTFKGTPTAENPEY 1248
DB 574 ESSPYWIGSNHQ-----INLDNPDY 594

RESULT 13
S35745
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus
C:Species: avian erythroblastosis virus
C:Date: 03-Mar-1994 #sequence_revision 26-May-1995 #text_change 28-Feb-1997
C:Accession: S35745
R:Vennstrom, B.
submitted to the EMBL Data Library, March 1993
A:Reference number: S35743
A:Accession: S35745
A:Molecule type: DNA
A:Residues: 1-544 <VEN>
A:Cross-references: EMBL:X12707
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific p
F:135-400/Domain: protein kinase homology <KIN>
F:143-151/Region: protein kinase ATP-binding motif
F:170/Active site: Lys #status predicted

Query Match 23.1%; Score 1575; DB 2; Length 544;
Best Local Similarity 53.4%; Pred. No. 4.4e-58;
Matches 336; Conservative 70; Mismatches 129; Indels 94; Gaps 17;

QY 578 GPEADQCACAHYKDPFCVACPSGVKPDLSYMPIWKFPDEBAGACQPCPINCTHSCVDL 637
DB 1 GP--DHCMKCAHFIDGPHCVKACPAVGLGENDTL-VMKYADANAVCQLCHPNCNCTRGKGP 57

QY 638 DDKGCAPAEQASPLTSIVSAV-GILLVVVLGVVFGILLIKRQOKIRKYIMRRLLQETEL 696
DB 58 GLEGCP---NGSKTPSTAAGVVGGLCLLVVGLGIGLYLRR-HIVAKRTLRRLQOREL 113

QY 697 VEPLTPSGAMPNOAMRILKEQYIKANSKFTIGTEL-TVYKGIWIPDGENVKIPVAIKVL 755
DB 114 VEPLTPSGEAPNOAHLILKETEFK-KVKVLGSAFGTYKGLWIPSEKVKIPVAIKEL 172

QY 756 RENTSPKANKEILDEAYMAGVGSYVSRLLGICLTSTVOLTMYPYGCGLLDHVRNRRG 815
DB 173 REATSPKANKEILDEAYMASVDNPHVCRLLGICLTSTVOLITQMPYGCGLLDYIREHKD 232

QY 816 RLGSODLLNWMQIAKMSYLEVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDET 875
DB 233 NIGSQYLLNWCQIAKGMNYLEERHVMVHRDLAARNVLKTPQHVKITDFGLAKQLGADEK 292

QY 876 EYHADGGKVPIKMALESILRRRFTTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDL 935
DB 293 EYHAGGKVPIKMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIPIASEISSV 352

QY 936 EKERLPQPPICTIDVYIMVWKMIIDSECRPRELVSEFSRMARDPQRFVVIQ-NEDL 994
DB 353 EKERLPQPPICTIDVYIMVWKMIIDADSRPKFRELIAEFKSMARDPPRYLVITQDERM 412

QY 995 GPASPLDSTYRSLLEDDMDLVDAEYLVPOGFFCPDPAFCAGGQVWHRHRSSTRS 1054
DB 413 HLPSPSTDSKYRFLMEBEDMEDIVDAEYLVPHQGF-----NSPST-- 454

QY 1055 GGGDLTLGLPSEEEAPRSP-----APSEGAGSDVDFDGLGMAAGLQSLPHTHDPSP 1109
DB 455 -----SRTPLLSLSATSNNSATNCIDRNGG-----H----- 481

QY 1110 QRYSEDPVPLPST-DGVYAPITCSPQBYVNPQVDRPQPPSPREGPLPAARPAAGAT-LE 1168
DB 482 -----PVREDGFL-----PAPEYVQ--LMPKKPSTAMVQNYVYISLTAIS 522

QY 1169 RAKTLSPKNGVVKVDFAFGGAIVENPEYL 1197
DB 523 KLPIDSRYQN-----SHSTAVDNPEYL 544

RESULT 14
S00727
kinase-related transforming protein (erbB) (EC 2.7.1.1) - avian erythroblastosis virus
C:Species: avian erythroblastosis virus
C:Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 28-Feb-1997
C:Accession: S00727
R:Scotting, P.; Vennstrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.
Oncogene Res. 1, 265-278, 1987
A:Title: Common site of mutation in the erbB gene of avian erythroblastosis virus mutant
A:Reference number: S00727; MUID:88217326; PMID:2897102
A:Accession: S00727
A:Molecule type: DNA
A:Residues: 1-545 <SCO>
A:Cross-references: EMBL:X06943
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; phosphotransferase
F:135-400/Domain: protein kinase homology <KIN>
F:143-151/Region: protein kinase ATP-binding motif

Query Match 23.0%; Score 1568; DB 2; Length 545;
Best Local Similarity 53.4%; Pred. No. 8.5e-58;

Matches 336; Conservative 69; Mismatches 130; Indels 94; Gaps 17;									
Qy	578	GPEADQCVACAHYKDPFCVACRPSGKPDLSYMPIMWKFDPDEGACQPCPINCTHSCVDL	637						
Db	1	GP--DHCWKCAHFDGPHCVKACPAVLGENDTL-VKMYADANAVCVCHPNCRTGCKGP	57						
Qy	638	DDKCPAEQASPLTSIVSAVV-GILLVVVLGVFGILLIKRQKQKIRYIMRRLLOTEL	696						
Db	58	GLECGP---NGSKTPSTAAGVVGGLCLLVVGLGIGLYLRR-HIVKRTLRLLQREL	113						
Qy	697	VEPLTPSGAMPNQAMRILKEQYIKANSKPGITEL-TVYGIWIPDGENVKIIPVAIKVL	755						
Db	114	VEPLTPSGEAPNQAHILKETEKF-KVKVLGFGAGTVYKGLWIPGEKVTIPVAIKEL	172						
Qy	756	RENTSPKANKEILDEAYMAGVGSPPYVSRLLGICLTSTVQLVTLQMPYGCILLDHVRENRG	815						
Db	173	REATSPKANKEILDEAYMASVDNPHVCRLLGICLTSTVQLITQMPYGCILLDVIREHKD	232						
Qy	816	RLGSODLLNMCOTAKMSVLEDRVLVHRDLAARNVLKSPNHVKITDFGLARLLDIDET	875						
Db	233	NIGSQYLLNMCVQAKGMNYLEERHLVHRDLAARNVLKTPQDVKITDFGLAKQLGADEK	292						
Qy	876	EYHADGKVPKIKWMALESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLL	935						
Db	293	EYHAGGKVPKIKWMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIPIASEISVL	352						
Qy	936	EKGERLPOPPCTIDVYIMVWKMDSECRPRELVSFBSRWARDPQRFVVIQ-NEDL	994						
Db	353	EKGERLPOPPCTIDVYIMVWKMSDADSRKPRELIAEFSKWARDPPRYVLVLOGDERM	412						
Qy	995	GPASPLDSTYRSLEDDMGDLVDAEYLVPOQGFPCPDAPAGAGGVHHRSSSTRS	1054						
Db	413	HLPSPTDSKPYRTLMEEDMEDIVDADEYLVPHQGF-----NSPST--	454						
Qy	1055	GGGDLTLGLPSEBEAPRSPL-----APSEGAGSDVFDGDLGMGAAGLQSLPHDPSPL	1109						
Db	455	-----SRTPLLSLSATSNNSTNCIDRNG-----H-----	481						
Qy	1110	QRYSEDTPTVLPSETDGYVAPLTCSPQPEYVQPDVVRPPSPREGPLPAARPAAGAT-LE	1168						
Db	482	-----PVREDGFL-----PAPEYVQ--LMPKKPSTAMVQNIYVLSLTAIS	522						
Qy	1169	RAKTLSPGKGVKVDVFAFGAVENPEYL	1197						
Db	523	KLPMDSRYQN-----SHSTAVDNPEYL	544						
RESULT 15									
QGFPE									
epidermal growth factor receptor - fruit fly (Drosophila melanogaster)									
N;Contains: protein-tyrosine kinase (BC 2.7.1.112) erBB									
C;Species: Drosophila melanogaster									
C;Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 11-Jun-1999									
C;Accession: A00640; A38021									
R;Livneh, B.; Glazer, L.; Segal, D.; Schlessinger, J.; Shilo, B.Z.									
Cell 40, 599-607, 1985									
A;Title: The Drosophila EGF receptor gene homolog: conservation of both hormone binding									
A;Reference number: A00640; MUID:85124611; PMID:2982499									
A;Accession: A00640									
A;Molecule type: DNA									
A;Residues: 1-1330 <liv>									
A;Cross-references: EMBL:K03054									
R;Wadsworth, S.C.; Vincent III, W.S.; Bilodeau-Wentworth, D.									
Nature 314, 178-180, 1985									
A;Title: A Drosophila genomic sequence with homology to human epidermal growth factor re									
A;Reference number: A38021; MUID:85137938; PMID:2983232									
A;Accession: A38021									
A;Molecule type: DNA									
A;Residues: 'A', 832-866, 'V', 868-943, 'QTPSLVK' <WAD>									
A;Cross-references: EMBL:X02293; NID:g7922; PIDN:CAA26157.1; PID:g929565									
C;Comment: This sequence is tentative because the introns have not been identified.									
C;Genetics:									
A;Gene: FlyBase:Bgfr									

A;Cross-references: FlyBase:FBgn0003731

A;Map position: 2 57P

C;Superfamily: epidermal growth factor receptor; protein kinase homology

C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho

F;1-732/Domain: extracellular #status predicted <EXT>

F;733-764/Domain: transmembrane #status predicted <TM>

F;765-1330/Domain: intracellular #status predicted <INT>

F;808-1072/Domain: protein kinase homology <KIN>

F;122-300.324.363.518.688.695.700/Binding site: carbohydrtate (Asn) (covalent) #status pr

F;774/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted

F;843/Active site: Lys #status predicted

F;1181/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status predict

Query Match 23.0%; Score 1565.5; DB 1; Length 1330;

Best Local Similarity 29.0%; Pred. No. 2.6e-57;

Matches 402; Conservative 182; Mismatches 417; Indels 385; Gaps 42;

Qy 80 VQGYVLIHQVQVPLQRLIRIVRGTLQF-----EDNYALAVLDNGDPLNTPPTVTGASP 134

Db 38 ITNYIVIGLDLPCTLSYRLQIRGRTLFSLSVEEEKYALFV-----TY 81

Qy 135 GGLREQLRLSLTEILKGGVLIQRNPOL-CYODTILWKDIFHKNQALTLTIDNRSRACHP 194

Db 82 SKMTTEIPDLRDLVNGQVGFHNNYNLCHMRTIOWSEIVNGTDAYNYNYDFTAPERBCPK 141

Qy 195 CSPMKGSRGWSSESDCQSLTRTVGAGCA--RCKGPLPTDCCHECQAAAGCTGPKHSDC 252

Db 142 CHESCTHG-CWGGSPKNCQKFSKLTCSQCAGRCYGPKECHLFCAGCTGTQKDC 200

Qy 253 LACLHNHSGICELHCPALVYNTDTFESMNPBEGRYTFQASCVTACPYNYLSTDVGSCT 312

Db 201 IACKNPFDEAVSKEECPPMRKYNPTTVLETNPEGYAYGATCVKECP-GHLLRDNGACV 259

Qy 313 LVCPHLNQEVTAEDGTQCEKSKPCARVCYGLGMQYIKANSKEFIGTEL-----EFAG 366

Db 260 RSCPQKMDKGE-----CVPCNGPCPKTC-----PGVTVLHAGNIDFRN 300

Qy 367 CKKIFGSLAPLPSFDG--DPASNTA-----PQPEQLQVFTLEITGYLISAMPDS 418

Db 301 CTVIDGNIRILDQTFSGFDVYATMGPRYIPLDPERREVFTVKETITGLYLNIEGTHPQ 360

Qy 419 LPDLSVFQNLQVIRGRIHNGAY-SLTQLGIGISWGLRSLRELGLSLALHNNHLCFV 477

Db 361 FRNLSYFRNLLETIHGRQLMESPAALAIKVSLSYLEMRNLKQISSGVVVIQHNRLDCYV 420

Qy 478 HTVPQDLFRNPHQALLHTANRPEDBC-----504

Db 421 SNIRWPAIQKEPEQKVVNENLRADLCGKFLTLISVQHNIIMHIFAICREKNHLLGSV 480

Qy 505 -----504

Db 481 QRGRLIGSWHGSVPYVQLQELQFQWHLHRLWLYIQVINSIQDKNSEHQLTDACYSPSVPT 540

Qy 505 -----VG 506

Db 541 SLTIERARVAIQSAGLAMELEQITARSASMRHSKTLPAEGRQVPRWVFLGVCSARAGIA 600

Qy 507 EGLA-----CHQLCARGHCWGPQTQVNCQFLRGQCEVCECVLQGLPREYV---N 556

Db 601 EPLAGRAVCRKTHPLCELCTNYGHEQVCSKCTHYKRREQCETEC-----PADHYTDEE 654

Qy 557 ARHCLPCHPECQPNQSVTCFGEAEOCVACAHYK-----DPPF-----CVARCPG 603

Db 655 QRECFQHFEC---NG---CTGGADDCSKCRNFKLPDANETGYVNSTWNTFNTSKCPLE 708

Qy 604 VK-PDLSYMPINKPPEDEGACQPCPINCTHSCVDLDDKGCQPARSPILTSIYSAVVGIL 662

Db 709 MRHVYQYTAIGPY-----CAASPPSSSKITANLD-----VNMFIITGAV 749

Qy 663 LVVVLGVVFGI-LIKRQKIRKYT--MRLLQETELVELPTFSGAMPNQAOQRILKEQY 719

Db 750 LVPTICILCVVYICRQKQKAKETVMTNLSGREDESEPLRPSNIGANLCKLRIVKDAE 809

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 09:10:23 ; Search time 22.179 Seconds
(without alignments)
5413.772 Million cell updates/sec

Title: SEQ4-710-730-12
Perfect score: 6775
Sequence: 1 MEAALCRWGLLLALLPPGA.....TFKGTPTAENPEYGLDVPV 1249
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 73.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6602	97.4	1255	1 A24571	protein-tyrosine k
2	5805	85.7	1260	1 TVRTNU	protein-tyrosine k
3	5795.5	85.5	1254	2 I48151	p-185 precursor -
4	3030	44.7	1210	1 GQHUE	epidermal growth f
5	3001	44.3	1210	2 A53183	epidermal growth f
6	2978.5	44.0	1223	1 TVCHLV	epidermal growth f
7	2849.5	42.1	1308	2 A47253	epidermal growth f
8	2553	37.7	1166	1 S06142	protein-tyrosine k
9	2305.5	34.0	1342	2 A36223	kinase-related tra
10	2218.5	32.7	1339	2 JC4387	epidermal growth f
11	1650.5	24.4	698	1 TVFVLV	protein-tyrosine k
12	1588	23.4	604	1 TVYUHV	protein-tyrosine k
13	1537	22.7	544	2 S35745	protein-tyrosine k
14	1530	22.6	545	2 S00727	kinase-related tra
15	1523.5	22.5	1330	1 GQFFE	epidermal growth f
16	1513	22.3	540	2 B47776	protein-tyrosine k
17	1511	22.3	540	1 TVFVFB	protein-tyrosine k
18	1509	22.3	644	2 A36325	epidermal growth f
19	1224	18.1	1323	2 B88257	protein let-23 [im
20	1224	18.1	1374	2 S70712	protein-tyrosine k
21	1139	16.8	1369	2 S70713	protein-tyrosine k
22	1126	16.6	527	2 A40232	epidermal growth f
23	1109	16.4	1717	1 A45558	epidermal growth f
24	975.5	14.4	843	2 A27131	epidermal growth f
25	806.5	11.9	346	2 S13807	protein-tyrosine k
26	754.5	11.1	311	2 S13808	protein-tyrosine k
27	703	10.4	1363	2 T43220	insulin-like growth
28	676.5	10.0	1372	2 A34157	insulin receptor p
29	675	10.0	1382	1 INHUR	insulin receptor p

ALIGNMENTS

RESULT 1

A24571
protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human
N/Alternate names: c-erb-B-2 protein precursor; kinase-related transforming protein erbB2
C/Species: Homo sapiens (man)
C/Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999
C/Accession: A24571; A25491; A44188; B44188; I59509; I57622
R/Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T.; Nature 319, 230-234, 1986
A/Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth factor
A/Reference number: A24571; MUID:86118663; PMID:3003577
A/Accession: A24571
A/Molecule type: mRNA
A/Residues: 1-1255 <YAM>
R/Cross-references: GB:X03363; NID:g31197; PIDN:CAA27060.1; PID:g31198
R/Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.
Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985
A/Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epidermal
A/Reference number: A25491; MUID:86016729; PMID:2995967
A/Accession: A25491
A/Molecule type: DNA
A/Residues: 737-1031 <SEM>
R/Cross-references: GB:M11767; NID:gi82163; PIDN:AAA35808.1; PID:g553282
R/Coussens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg, P.
Science 230, 1132-1139, 1985
A/Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chromosomal
A/Reference number: A44188; MUID:86070181; PMID:2999974
A/Accession: A44188
A/Molecule type: DNA
A/Residues: 740-910 <COU1>
R/Cross-references: GB:M12036; NID:gi83988; PIDN:AAA35978.1; PID:g183989
A/Accession: B44188
A/Molecule type: mRNA
A/Residues: 1-517, 'RALL', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>
R/Cross-references: GB:M11730; NID:gi83986
R/King, C.R.; Kraus, M.H.; Aaronson, S.A.
Science 229, 974-976, 1985
A/Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.
A/Reference number: I59509; MUID:85272597; PMID:2992089
A/Accession: I59509
A/Status: translated from GB/EMBL/DDBJ
A/Molecule type: DNA
A/Residues: 832-909 <REX>
R/Cross-references: GB:I29395; NID:g459807; PIDN:AAA35809.1; PID:g459808
R/Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.
Mol. Cell. Biol. 7, 2597-2601, 1987
A/Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptional
A/Reference number: I57622; MUID:87286898; PMID:3039351
A/Accession: I57622
A/Status: translated from GB/EMBL/DDBJ
A/Molecule type: DNA
A/Residues: 1-191 <TAL>

insulin receptor p
insulin receptor-r
protein-tyrosine k
insulin-like growth
insulin receptor-r
insulin-like growth
insulin-like growth
protein-tyrosine k
protein-tyrosine k
insulin receptor -
protein-tyrosine k
insulin receptor (-
protein-tyrosine k
insulin receptor -
mouse developmenta
eph-related recept

A;Cross-references: GB:M16792; NID:g183983; PIDN:AAA58637.1; PID:g553332
 C;Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
 C;Genetics:
 A;Gene: GDB:ERBB2; NGU: NEU; HER-2
 A;Cross-references: GDB:120613; OMIM:164870
 A;Map position: 17q21.1-17q21.1
 A;Introns: 25/1; 75/3; 147/1; 883/3
 A;Note: the list of introns is incomplete
 C;Function:
 A;Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
 C;Superfamily: epidermal growth factor receptor; protein kinase homology
 C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
 inase
 F;1-21/Domain: signal sequence #status Predicted <SIG>
 F;22-1255/Product: protein-tyrosine kinase erbB2 #status Predicted <MAT>
 F;22-653/Domain: extracellular #status Predicted <EXT>
 F;70-304/Domain: EGF receptor extracellular domain repeat <EB1>
 F;395-605/Domain: EGF receptor extracellular domain repeat <BE2>
 F;654-675/Domain: transmembrane #status Predicted <TMM>
 F;676-1255/Domain: intracellular #status Predicted <INT>
 F;718-983/Domain: intracellular #status Predicted <KIN>
 F;726-734/Region: protein kinase homology <KIN>
 F;68,124,187,259,530,571,629/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F;686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
 F;753/Active site: Lys #status predicted
 F;1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 97.4%; Score 6602; DB 1; Length 1255;
 Best Local Similarity 97.2%; Pred. No. 3.2e-277;
 Matches 1221; Conservative 6; Mismatches 21; Indels 8; Gaps 2;

Qy 1 MELAALCRWGLLLALLPPGAASCTGCTDMKRLPASPETHDMLRHLVGGCQVVOGNL 60
 Db 1 MELAALCRWGLLLALLPPGAASCTGCTDMKRLPASPETHDMLRHLVGGCQVVOGNL 60

Qy 61 ELTYLPTNASLSFLQDIQEVGYVLIHNRVQVPLQRLRVRGTLFEDNYALAVLDNG 120
 Db 61 ELTYLPTNASLSFLQDIQEVGYVLIHNRVQVPLQRLRVRGTLFEDNYALAVLDNG 120

Qy 121 DPLNTTTPVTGASPGGLREQLRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHNKOLA 180
 Db 121 DPLNTTTPVTGASPGGLREQLRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHNKOLA 180

Qy 181 LTLDTNRSRACHPCSPMKSGSRGWGSSSDQSLTRTVACGGCARCKGPLPTDCHEQC 240
 Db 181 LTLDTNRSRACHPCSPMKSGSRGWGSSSDQSLTRTVACGGCARCKGPLPTDCHEQC 240

Qy 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy 301 YNYLSTDVGSCTLYCPLHNQVTAEDGTQCEKSKPCARVCYGLGHQYIKANSKFTGIT 360
 Db 301 YNYLSTDVGSCTLYCPLHNQVTAEDGTQCEKSKPCARVCYGLGHQYIKANSKFTGIT 360

Qy 361 ELEFAGCKIFGSLAFIPESFDGDPASNTAPLOEQLOVPELTBEITGYLYISAWPDSLP 420
 Db 361 IQEFAGCKIFGSLAFIPESFDGDPASNTAPLOEQLOVPELTBEITGYLYISAWPDSLP 420

Qy 421 DLSVFQNLQVIRGRILHNGAYSLTQGLIGISWLGLRSLRGLSGLALIHNTLHLCFVHTV 480
 Db 421 DLSVFQNLQVIRGRILHNGAYSLTQGLIGISWLGLRSLRGLSGLALIHNTLHLCFVHTV 480

Qy 481 PWDOLFNPQHALLHTANRDECEVCGELACHQICARGHCWGPGTQCVCNSQFLRQEC 540
 Db 481 PWDOLFNPQHALLHTANRDECEVCGELACHQICARGHCWGPGTQCVCNSQFLRQEC 540

Qy 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFPGPADQCVACAHYKDPFCVARC 600
 Db 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFPGPADQCVACAHYKDPFCVARC 600

Qy 601 PSGVKPDLSPYIWKFPDEEGACQPCPINTCHSCVDLDDKGCQPAEQRASPLTSTISAVVG 660

Db 601 PSGVKPDLSPYIWKFPDEEGACQPCPINTCHSCVDLDDKGCQPAEQRASPLTSTISAVVG 660
 Qy 661 ILVWVVGVFGLIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKTEL 720
 Db 661 ILVWVVGVFGLIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKTEL 720
 Qy 721 RKVKVLGSGAFGQYIKA-----NSKFIGITELVIRENTSPKANKEILDEAYVMAGVGS 773
 Db 721 RKVKVLGSGAFGTVYGIWIPDGENVK-IPVAIKVIRENTSPKANKEILDEAYVMAGVGS 779
 Qy 774 PYVSRLIGICLTSTVQLVTQLMYGCLLDHVRENRGLSGQDLLNWCQIAGKMSYLEDV 833
 Db 780 PYVSRLIGICLTSTVQLVTQLMYGCLLDHVRENRGLSGQDLLNWCQIAGKMSYLEDV 839
 Qy 834 RLVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKMMALRESILRRRF 893
 Db 840 RLVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKMMALRESILRRRF 899
 Qy 894 THQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVKCW 953
 Db 900 THQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVKCW 959
 Qy 954 MIDSECRPRPRELVSEFSRWARDPQRPVJTQNEIDLGPASPLDSTFYRSLLDEDDMGDLVD 1013
 Db 960 MIDSECRPRPRELVSEFSRWARDPQRPVJTQNEIDLGPASPLDSTFYRSLLDEDDMGDLVD 1019
 Qy 1014 AEYLVPOQGFCCPDPAAGAGGMVHRHRSSTRSGGDLTLGLEPSEEEAPRSLAPSE 1073
 Db 1020 AEYLVPOQGFCCPDPAAGAGGMVHRHRSSTRSGGDLTLGLEPSEEEAPRSLAPSE 1079
 Qy 1074 GAGSDVFDGDLGMAAKGLQSLTFHDPSPLOQRYSEDPVPLPSETDGYVAPLTCSPQPEY 1133
 Db 1080 GAGSDVFDGDLGMAAKGLQSLTFHDPSPLOQRYSEDPVPLPSETDGYVAPLTCSPQPEY 1139
 Qy 1134 VNQPDVVRPQPPSPREGPLPAARPAATLERAKTILSPGKGVKVDVFAFGGAVENPEYLTTP 1193
 Db 1140 VNQPDVVRPQPPSPREGPLPAARPAATLERAKTILSPGKGVKVDVFAFGGAVENPEYLTTP 1199
 Qy 1194 QGAAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVPV 1249
 Db 1200 QGAAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVPV 1255

RESULT 2

TVRTNU

protein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 11-Jun-1999

C;Accession: A24562; A61204

R;Bargmann, C.I.; Hung, M.C.; Weinberg, R.A.

Nature 319, 226-230, 1986

A;Title: The neu oncogene encodes an epidermal growth factor receptor-related protein.

A;Reference number: A24562; MUID:86118662; PMID:3945311

A;Accession: A24562

A;Molecule type: mRNA

A;Residues: 1-1260 <BAR>

A;Cross-references: EMBL:X03362; NID:g56745; PIDN:CAA27059.1; PID:g56746

R;Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohen,

Carcinogenesis 12, 1975-1978, 1991

A;Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals no m

2-thiazolyl]formamide or N-methyl-N-nitrosourea.

A;Reference number: A61204; MUID:92035293; PMID:1682063

A;Accession: A61204

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 637-663,'V',665-702 <MAS>

A;Note: authors translated the codon GCA for residue 25 as Val

C;Genetics:

A;Gene: neu

C;Superfamily: epidermal growth factor receptor; protein kinase homology

C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho

F;1-19/Domain: signal sequence #status predicted <SIG>

F;20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>

A;Molecule type: mRNA
A;Residues: 969-971,'K',973-1115,'D' <EIS>
A;Cross-references: EMBL:Z12608
R;Heisermann, G.J.; Gill, G.N.
J. Biol. Chem. 263, 13152-13158, 1988
A;Title: Epidermal growth factor receptor threonine and serine residues phosphorylated
A;Reference number: A28941; MUID:88330814; PMID:3138233
A;Accession: A28941
A;Molecule type: protein
A;Residues: 689-694,'X',696-704,'L',706-707,989-992,'XX',995-996,'X',998-1000;1002-1009,
R;Hibbs, M.L.; Dunn, A.R.; Alexander, W.S.
submitted to the EMBL Data Library, April 1994
A;Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor
A;Reference number: S45325
A;Accession: S45325
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-971,'K',973-1210 <VER>
A;Cross-references: EMBL:X78987; NID:G488830; PIDN:CAA55587.1; PID:G488831
R;Paria, B.C.; Das, S.K.; Andrews, G.K.; Dev, S.K.
Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993
A;Title: Expression of the epidermal growth factor receptor gene is regulated in mouse b
A;Reference number: I49643; MUID:93126380; PMID:7678348
A;Accession: I49643
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 12-20,132 <RES>
A;Cross-references: GB:I06864; NID:G193001; PIDN:AAAS3029.1; PID:G567201
C;Genetics:
A;Gene: EGFR
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphop
F;1-24/Domain: signal sequence #status predicted <SIG>
F;648-670/Domain: transmembrane #status predicted <TM>
F;712-977/Domain: protein kinase homology <KIN>
F;680,695/Binding site: protein kinase ATP-binding motif
F;720-728/Region: site: phosphate (Thr) (covalent) #status experimental
F;697,1070,1071/Binding site: phosphate (Ser) (covalent) #status experimental
F;993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental
F;1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental
F;1197/Binding site: phosphate (Tyr) (covalent) #status experimental

Query Match 44.3%; Score 3001; DB 2; Length 1210;
Best Local Similarity 48.6%; Pred. No. 3.8e-122;
Matches 620; Conservative 168; Mismatches 365; Indels 122; Gaps 28;

QY 11 LLALALLPGAA--STQVCTGTDKMLRLPASPETHLDMRLHYOCQVQVQGNLELTYPTN 68
DB 14 LLTALCAAGALBEKKYCCQTSNRLTQGTGFEDHFLSLQRMYNCEVTLGNLEITYVQRN 73

QY 69 ASLSFLQDIOEVGYVLIANQVROVPLORLRIVRGTQOLFEDNYVALVLDNGDPLNNTTP 128
DB 74 YDLSFLKTIQEVAGYVLIANTVERIPLENLQIIRGNALYENTYVALISN----- 124

QY 129 VTGASPGGLRELQRLSITELKGGVLTQRNPOLCYODTLWKDI----FHKNNQLAULTLI 184
DB 125 -YGNTRTGLRELPMNLQELIIGAVRFSNNPILCNMDTIQWRDI VQNVFNSWMDL--- 180

QY 185 DTVRSRACHPCSPMCKGSRGWGSSSEDCSLTRTVVAGGCA-RCKGPLPTDCHEQCAAG 243
DB 181 -QSHSSCPKCDPSCPNCSNGSCWGGEENCQKLTIIQAQCSHRCRCGRSPSDCHNCAAG 239

QY 244 CTGPKHSDDLACLFHNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCYTACPNY 303
DB 240 CTGPRESDDLVCOKFQDEATCKDTCPLMLYNTPTTQMVNPEKYSFGATCVKKCPRNY 299

QY 304 LSTDVGSCTLVCLPHNQEVTAEDGTORCEKSCPKARVCYGLQWQIKANSKFIGITELE 363
DB 300 VVTDHSCVRACGPDYEV-EEDGIRKCKCDGCRKVCNGIGIGEFK-DTLSINATNIK 357

QY 364 -FAGCKIFGSLAFLESFPGDPSANTAPLQPSQLQVFTLEBITGYLYISAWPDSLPDL 422
DB 358 HFKYCTAISGDLHILPVAFKGDSFTRTPPLDPRLEILTKVKEITGFLTAQAPDNWTDL 417

RESULT 6 TVCHLV

N;Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
C;Species: Gallus gallus (chicken)
C;Date: 28-Feb-1986 #sequence_revision 05-May-1995 #text_change 04-Feb-2000
C;Accession: A27720; A00643
R;Lax, I.; Johnson, A.; Howk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Vennstr
Mol. Cell. Biol. 8, 1970-1978, 1988
A;Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mou
A;Reference number: A27720; MUID:88261272; PMID:3260329

QY 423 SVFQNLQVIRGRILHNGAYSLTQGLGISWLGRLSRLSGSLALIHNTLHLCFVHTVPW 482
DB 418 HAFENLEIRGRTRKQGFSLAVGLNITSLGRLSKEISDGVIIISGRNLCYANTINW 477

QY 483 DQLFNPHQALLHTANRDECEVGEGLAQHQLCARGHCWGPQTQCVCNSQFIRGQECVE 542
DB 478 KKLFGTPNQTKIMNNAEKDKCAVNHVCNPLCCSSEGCWGPEDPCVSCQNVSRGECVE 537

QY 543 ECEVLQGLPREYVNAHCHLPCHPECOPONGSVTCFGEADQCACAHYKDPPECVARCPS 602
DB 538 KCNILEGBREFEVENSECQCHPECLPQANNTCTGRGPDNCIQCAHYIDGPHCVKTCPA 597

QY 603 GVKPDLSPYPIWKFPEDEGACQPCNCTHSCVDLDDKGPABQARASPLTISVAVVGIL 662
DB 598 GINGENNTL-VMKYADANNVCHLCHANCTYGCAGPGLQCEVWPSPKPISTIATIGVGL 656

QY 663 LVVLGVVFGI-LIKERQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRLIKETELR 721
DB 657 LFTIV-VALGIGLFMRRRHIVRKRTLRLQLQERELVEPLTPSGEAPNQAHRLIKETEFK 715

QY 722 KYVLGSGAFGQ-----YIKANSKF---IGITELVIRENTSPKANKEIIDEAYVMAGVS 773
DB 716 KIKVLGSGAFGTYYKGLWPEGEKVIPVAIKE--LREATSPKANKEIIDEAYVMASVDN 773

QY 774 PYVSRLLGICLTSTVOLTPYGLLDHVRNRCRLGSDLLNWCMIQAKGMSYLEDV 833
DB 774 PHVCRLLGICLTSTVOLITQPLPYGCLLDYVREHKDNIGSYLLNWCVIQAKGMVLEDR 833

QY 834 RLVRDLAARNVLKSPNPKITDFGLARLLDIDETEHADGKVPKIMMALESILRRRF 893
DB 834 RLVRDLAARNVLKTPQHKITDFGLAKLGAEEKEYHAEGKVPKIMMALESILHRIY 893

QY 894 THQSDVMSYGVTVWELMTGAKPYDGPARELPDLLEKGERLPDPPICTIDVYMIWVKCW 953
DB 894 THQSDVMSYGVTVWELMTGSKPYDGPASDISILEKGERLPDPPICTIDVYMIWVKCW 953

QY 954 MIDSECRPFRELVSFSSRMARDPQRFVIO-NEDLGASPLDSTFYRSLLEDDMDGLV 1012
DB 954 MIDADSRPFRELILFSSQWARDPQRYLVIOGDERHLPSPDTSNFYRALMDEDMEDVV 1013

QY 1013 DAEYLVPQOGFPCDPAPGAGMWHHRSSSTRSGGDLTLGLPEPSEEAAPRSLAPS 1072
DB 1014 DADEYLVPQOGF-----NSPST-----SRTLSS 1039

QY 1073 EGNQSDVFDGDLGMGAAGKLSLPHDPSLPQRYSESDTPVLPSET--DGYVAPLTCSPQ 1130
DB 1040 LSATSN---NSTVACINRNGSCRVKEDAFQRYSSDPTGCAVTEIDIDDAFL-----PV 1089

QY 1131 PEYVNPQDVPPOPSPREGPLPAARPAAGATLERAKTLPCKNGVVKDVPFAGGAVENPEY 1190
DB 1090 PEYVNO-SYPKPAGSVQNPVYHNQPLHP-----AFGRDLHYQN--PHSNVAGNPEY 1138

QY 1191 L-TPQGAAPHPHPPPAFPDNLVYWDQ-----DP-----PERGAPPSTFKG 1233
DB 1139 LNTAQ-----PTCLSSGNSPALWIKGSHQMSLDNDPDYQODFFPKETKPNKIFKG 1189

QY 1234 TPTAENPEYLGLDVP 1248
DB 1190 -PTAENAEYLRVAPP 1203

A;Accession: A27720
A;Molecule type: mRNA
A;Residues: 1-1223 <LAX>
A;Cross-references: GB:M20386
R;Nilsson, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1995
A;Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A;Reference number: A00643; MUID:95228222; PMID:2988784
A;Accession: A00643
A;Molecule type: mRNA
A;Residues: 585-1223 <NIL>
A;Cross-references: GB:M10066
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
specific protein kinase
F;1-30/Domain: signal sequence #status predicted <SIG>
F;31-1223/Product: epidermal growth factor receptor #status predicted <MAT>
F;31-654/Domain: extracellular #status predicted <EXT>
F;81-307/Domain: EGF receptor extracellular domain repeat <EE1>
F;397-610/Domain: EGF receptor extracellular domain repeat <EE2>
F;655-677/Domain: transmembrane #status predicted <TM>
F;678-1223/Domain: intracellular #status predicted <INT>
F;719-984/Domain: protein kinase homology <KIN>
F;727-735/Region: protein kinase ATP-binding motif
F;136,202,280,361,370,422,575,580,615,635/Binding site: carboxydrate (Thr) (covalent) #
F;192,650/Binding site: carboxydrate (Ser) (covalent) #status predicted
F;687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F;754/Active site: lys #status predicted
F;1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat
Query Match 44.0%; Score 2978.5; DB 1; Length 1223;
Best Local Similarity 47.5%; Pred. No. 3.6e-121;
Matches 618; Conservative 173; Mismatches 352; Indels 157; Gaps 30;
QY 8 RWGLLLALLPPGAA-----STOVCTGTDMLRLPASPETHDMLRHLRYGCGVQVGNLE 61
DB 13 RGAALVLLLLGVALCSAVEBEKVCQCTNNKLTQLGHVEDHFTSLQRMVNNCEVLSNLE 72
QY 62 LTYLPTNASIFLODIOEVGYVLIHNOVROVLOLRIVRGTLFEDNYALAVLDNGD 121
DB 73 ITVEHNRDLTFUKTIOEVAGYVLIHNMVDVPLENLQIIRGNVLYDINSFALAVLSNYH 132
QY 122 PLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHKNQAL 181
DB 133 -NNKTO-----GLRELPMKELSELINGVVISNNPKLNCMDTVLWLDIITSRK-PL 182
QY 182 TLID-TNRSRACHPCSPMKSGRCWGESSEDCOSLRTVCAGCA-RCKGPLPTDCHEQ 239
DB 183 TVLDFASNLSSCPKCHPNCTEDHCWAGEQNCOTLTAKVCAQCSCRCGRKVPSCCHNQ 242
QY 240 CAAGCTGPKHSDCLACLNHNSGICELHCPALTYNTDTPESMPNPEGRVTCGASCVTAC 299
DB 243 CAAGCTGPRSDCLACKFRDDATCKDTCPLLYNPTTYQMDVNPGEKYSFGATCVRCE 302
QY 300 PYNLSTDVGSCITVCPPLHNOEYTAEDGTORCEKCKPCARVCYGLGMQVYKANSKFIGI 359
DB 303 PHNVVTDHGSCVRSNTDITYEV-BENGVRKCKCDGLSKVCNGIGELKGLS-INA 360
QY 360 TELE-FAGCKKIFGSLAFIPESFDGDPASNTAPLOPEQLQVFTLEITGYLISAMPDS 418
DB 361 TNIDSPKNCCTKINGDVSILPVAFGLDAFTKTLPLDKPLDVFTVKEISGFLLIQAWPDN 420
QY 419 LPDLSVFONLOVTRGILHNGAYSITLQGLISWLGRLSRLGSLALIHHTHLCFVH 478
DB 421 ATDLYAFENLEIIRGTQKQGYSLAVNLKIOSLGRSLKETSDDGAIAMKNKLCYAD 480
QY 479 TVPDOLFRNPQALLHTANRPEDECVGEGLACHQLCARGHCGPPTQCVNCSQFLRGQ 538
DB 481 TNNWSLFPATOSQTKLIQNRKNKNDADRHVCDPLCSGVCGMPGPFHCFCSFRFSRQK 540
QY 539 ECVECRVLQGLPREYVNAHCLPCHPECPQNG---SVTCFGPEADQCACAHYKDPFF 595

DB 541 ECVKQCNILQGEPRERDSKCLPCHSECLVQNSTAYNTTCSGPDHCKCAHFIDGPH 600
QY 596 CVARCPGKVPDLVYMPINKFPDEECACOPCPINCTHSCVDLDDKCPABQASPLTSIV 655
DB 601 CVACAPAGVLGENDTL-VWKYADANAVCOLCHPCNCTRGCKGPGLEGP---NGSKTPSIA 656
QY 656 SAVV-GILLVVLGVVFGILIKRQOKIRKYNNRLLQETELVEPLTPSGAMPNQAMRI 714
DB 657 AGVVGGLLCLVVLGVLGLVLRRL-HVVRKTLRRLQLERELVEPLTPSGEAPNQAHRI 715
QY 715 LKTELKRVKVLGSGAFQ-----YTKANSKP---IGITELVLRNTSPKANKEILDEAY 766
DB 716 LKETEKKVKVLGSGAFGVYKGLWPEGEKVIPAIRE--LREATSPKANKEILDEAY 773
QY 767 VMAGVSPVSRLLGLCLTSTVQLMPYGLLDHVRNCRGLSGQDLNNCMQIAGK 826
DB 774 VMASVNPVHVRLLGLCLTSTVQLMPYGLLDYIIRHKONIGSOYLLNNCVCQIAGK 833
QY 827 MSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMWALE 886
DB 834 MNYLEERLVRDLAARNVLKTPQHVKITDFGLAKLLGADEKEYHAEGGKVPKMWALE 893
QY 887 SILRRFTQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLLEKGERLPPOPICTIDVY 946
DB 894 SILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIIPASEISSVLEKGERLPPOPICTIDVY 953
QY 947 MIMVKCMMIDSECRPRFRELVSFSESRMARDPQRFVVIQ-NEDLGPASPLDSTFYRSILLED 1005
DB 954 MIMVKCMMIDADSRKPRFELIAEFKWARDPPRYLVIQGERMHLPSPTDSKEYFRTLMEE 1013
QY 1006 DMGDLDVDAEYLVPOQGFPCDPAPGAGGMVHRRHRSSTRSGGDLTLGLPSEBEAP 1065
DB 1014 EDMEDIVDAEYLVPHQGF-----NSPST-----S 1039
QY 1066 RSPL-----APBEGAGSDVFDGLGMAAGKQSLTHDPSPLQRYSEPTVPLPSET-- 1118
DB 1040 RTPLLSSLSATSNNSATNCID-----RNGQGHVREDSFQRYSSDPTGNFLESID 1091
QY 1119 DGVVAPLTCSPQEVYNQDVRPQPSREGPLPAARPAGATLERAKTLPSPGNVGVKDV 1178
DB 1092 DGLF-----FAPEYVNO--LMPKRP-----TAMVQONLYNNI 1122
QY 1179 F-----AFGAVENPEYLTPOGGAAPQHPPPAFSPAFNLYYWDQ---- 1219
DB 1123 SLTAISKLPWDSYQNSHSTAVDNPEYL-----NTQSPKLTQVTFESSPTWIOSGNH 1174
QY 1220 ----DPPE-----RGAPPSTFKGPTPTAENPEYLGLDVP 1248
DB 1175 QINLNDPDIQQOFLPNETKPNGLLKVPAAENPEYLRVAAP 1214
RESULT 7
A47253
Epidermal growth factor receptor, HER4 - human
C;Species: Homo sapiens (man)
C;Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C;Accession: A47253
R;Floman, G.D.; Culouscou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.; Ne
Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993
A;Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epidermal
A;Reference number: A47253; MUID:93189574; PMID:8383326
A;Accession: A47253
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: nucleic acid
A;Residues: 1-1308 <PLO>
A;Cross-references: GB:L07868; NID:g337359; PIDN:AAB59446.1; PID:g337360
A;Note: sequence extracted from NCBI backbone (NCBI:126842)
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; growth factor receptor
F;716-981/Domain: protein kinase homology <KIN>
F;724-732/Region: protein kinase ATP-binding motif

Query Match	42.1%	Score 2849.5	DB 2	Length 1308
Best local Similarity	44.1%	Pred. No. 1.4e-115		
Matches 593	Conservative 184	Mismatches 330	Indels 179	Gaps 29
Qy	9	WGULLALLPPGAA---STVCVCTGDMKRLRASPETHLDMLRHLYOGCVOVGNLELT	64	
Db	8	WVWSLLVAAGTVPDSQSVCAGTENKLSLSLEQOYRALRKYENCEVWGNLEITS	67	
Qy	65	LPTNASLSFLQDIOEVGYVLIAHNQVRQVPLQRLRIVRGQTQLPEDNYALAVLDNGDPLN	124	
Db	68	IEHRDLSFLRSVREVTGYVLVALNQFRYLPLENLRIRGTKLVEDRYALAIFLNYRKDG	127	
Qy	125	NTTPVTGASPGGLRELOLRSLTEILKGGVLIORNPQLCYOQTILWKDIFHKNQLALT	184	
Db	128	NF-----GLQELGKLNLTIELNGGVVVDQNKFLCYADTIHWQDIVENPWSNLT	178	
Qy	185	DTNRSRACHPCSPMCKSRGWSSSDCOSLTRITVCAGGC-ARCKGPLPTDCCHEQCAAG	243	
Db	179	STNGSSCGRCHKSCCTG-RWGPPTENHCQTLTRITVCAEQDCGRCYGVYSDCCCHRECAGG	237	
Qy	244	CTGPKHSDCLACLHFNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACPNY	303	
Db	238	CSGPKDTCFACMNFDSGACVTCQPTQFVYNPTTFQLEHFNKAYTYTGAFCKVCKEHN	297	
Qy	304	LSTDVSGCTILVCPLHNQEVTAEDGTQCEKCKSPCARVCYGLGMQYIKANSKFIGITE	363	
Db	298	V-VDSSCVACRSPSSKMEV-EENGIKMCKPCTDICPKACDGIPTGSLMSAQTVDDSSNIDK	355	
Qy	364	FAGCKKIFGSLAFPEFGDGPASNTAPLOEQLOVFEETLEETCYLIYISAWPDSLPDLS	423	
Db	356	FINCTKINGNLIFLVTGIGHGDPYNAIBADPEKLNVERTVREITGFLINIOSWPNMTDFS	415	
Qy	424	VFQNLQVIRGRILHNGAYSITLQGLGISWLSRLSRELGLSLAIHNHHLCFVHTVPWD	483	
Db	416	VFSNLVTIGRVLXSGLSLILKQOGLTSLOFQSLKEISAGNYITDNSNLCYHITHNT	475	
Qy	484	QLFRNPQOALLHTANRPEDECVGEGLAACHOLCARGHCWGPPTQCVNCSQFLRGQECVEE	543	
Db	476	TLFSTINQRIVRDNRKAENCTAEGMVCNHLCSDDGCGWGPQDCLSCRRFSRGRICIES	535	
Qy	544	CRVLQGLPRYVNAHCLCPHCECOP-ONGSVTCTFGPEADQCACAHYKDPDFCVACRPS	602	
Db	536	CNLYDGEFREFNGSICVECDPOCEKMEGDLTCHGFGPDNCTKCSHFKDGPNCKECPD	595	
Qy	603	GVKPDLSYMPIWKFDPDEGACQPCINCTHSCVDLDDKGC-----PAAEQASPL	651	
Db	596	GLOGANSF--IFKYADPRCHPCPNCTOGCNGPSTHDCIYYPWTGHSTLPQHAR-TPL	652	
Qy	652	TSIVSAVV-GILLVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQ	710	
Db	653	--IAAGVIGLFLIVIVGLTFAVYVRKSIK-KKRALRRFL-ETELVEPLTPSGTAPNQ	708	
Qy	711	QMRLKETELRUKVVLGSGAFQGVYK-----ANSKFIGITELVLRNTSPKANKEILDE	764	
Db	709	QLRLKETELRUKVVLGSGAFGVYKGIWVPEGETVKIPVAIKILNETTGPKANVEFMD	768	
Qy	765	AYVMAGVSPVSRLLGICLTSTVOLVTQLMPYGCILLDHYRNRGLSGODLLNWCWOIA	824	
Db	769	ALTMASMDHPLVRLGLVCUSPTQLVTLQMPHGCLLEYVHEHKDNIGSOLLNWCVOIA	828	
Qy	825	KGMSYLEDLVRLDRAARNVLKSPNHVKITDPLGLARLLDIDETEYHADGGKVPKWWA	884	
Db	829	KGMWYLEERLVRDLAARNVLKSPNHVKITDPLGLARLLEGEDEKEYNADGGKVPKWWA	888	
Qy	885	LESILRRFTHQSDVWSYGVTVWELMTFGAKPYDGDIPARIPLLLEKGERLPQPPCTID	944	
Db	889	LECITHYRKFTHQSDVWSYGVTVIWEMLTFGKPYDGTREIPDLLEKGERLPQPPCTID	948	
Qy	945	VYIMWKCWMIIDSECRPRFRELVEFSRMDARDPQRFVVIQNE-DLGPASPLDSTFYRSL	1003	
Db	949	VYIMWKCWMIIDADSRPKFKELAAEFSRMDARDPQRYLYIVQGDHMKLPSPNDSKFFQNLL	1008	
Qy	1004	EDDDMGDLVDAEEYLVPOQQGFFCPDPAPGAGGMVHRRHSRSTSRSGGDLTLGLEPSEB	1063	

[illegible]

Db 123 YQK-NPSSP--DVYQVGLKQLQLSNLTLISGGVKVSHNELLNVTINWMDIVDKTSNP 179

QY 180 ALTLIDNRSRACHPCSPMKGRWGSESSDCQSLTRITVCAGCG-ARCKGPLPTDCCHE 238

Db 180 TMLNLIHAFERQCKQKDHGCVNGSWAPGPGHCKQFTKLLCAEQCNRCRGPXPIDCCNE 239

QY 239 QCAAGCTGPKHSDCLACHHNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTA 298

Db 240 HCAGGCTGPRATCLACRFDNDDGTCKDTPPKPIYDIVSHVQVVDNPNIKYTFGAACVKE 299

QY 299 CPYNYLSTDVGSCTLVCPHLNQEVTAEDGFORCEKCKSPCARVCYGLGMQYIKANSKF 358

Db 300 CPNRYVUTE-GACVRSAGMLEVD-ENGRSKCPDGVCPKCDGIGISL-SNTIAVN 356

QY 359 ITL-BFAGCKKIFGSLAFELPESFDGDPASNTAPLOPEQLQVFETLEEITGYLYISAWPD 417

Db 357 STNIRSFNCTKINGDIILNRNSFEGDPHYKIGITMDPEHLWNLTTVKEITGYLVIMWPE 416

QY 418 SLPLSVFQNLQVIRGILHNGAYS-LTQGLGISWGLSLRELASGLALIHNNHLCF 476

Db 417 NMTSLSVFQNLQVIRGILHNGAYS-LTQGLGISWGLSLRELASGLALIHNNHLCF 476

QY 477 VHTVPWDQLFRNPHOALLHANPEDECVCGEGLACHOLCARGHCWGPQTCCVNCQSOFLR 536

Db 477 ANTNWRLEFSDQSIYDART-----ENQTCNNECEDGCV-PGPTMVCSLHVD 528

QY 537 GQCEVBEVRVLOGLPREYNARHCLPCHPECPQNGSVTCFGEADQCACAHYKDPFPC 596

Db 529 GGRCVASCNLLQGEPREAQVDGRCVQCHQECVQTSLSLTCYGPANCKSAHFQDGPQC 588

QY 597 VARPCEGVRPDLSYMPIWPKFDEGACQPCPINCTHSCVDLDDKGPAPORASPLTSIVS 656

Db 589 IPRCPHGILGDGTL-IWKYADRMQCCQPCQCHQCTQCGSGPLSGCRGD-IVSHSLAVG 646

QY 657 AVVGILLVVLGVVFGILIKRRQKRYKYMRLLOTELVEPLTPSGAMPNQAQRIK 716

Db 647 LVSGLLITVIALVLLVLLRRRIK-RKRTIRCLQELVELVEPLTPSGQAPNFAIRLIK 705

QY 717 ETELKRVKVLGSGAFQYIK-----ANSKFIGITELVIRENTSPKANKEILDEAYVMAG 770

Db 706 ETEFKDRVLGSGAFGVYKGLWNPGENIRIPVAIKVLREATSPKVNQEVLEAYVMAS 765

QY 771 VGSPPYVSRLLIGICLTSTVQLVTQLMYPGCLLDHVRNRRGLSGQDILLNWCQIAKMSYL 830

Db 766 VDPHPVCRLLIGICLTSAVQLVTQLMYPGCLLDVVRQHOERICQWLLNWCQIAKMSYL 825

QY 831 EDVRLVHRDLAARNVLKSNHVKITDGLARLLDIDETEHADGKGVPIKMALESILR 890

Db 826 EERHLVHRDLAARNVLLKNPNHVKITDGLSKULLTADKEKEYQADGGKVPDKMALESILQ 885

QY 891 RRETHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGRLLPOPPICITDVYMIW 950

Db 886 WTYTHQSDVMSYGVTVWELMTFGSKPYDGIIPAKEIASVLENGERLLPOPPICITVEMIL 945

QY 951 KCMWIDSECRPRELVSFESRMARDQPRVVTIONEDLGPASPLDSTFYRSLLDDEDMGD 1010

Db 946 KCMWIDPSRRPRELVSFESQMARPSRYLVIQG---NLPSSLRDLFRLLSSDD-D 1000

QY 1011 LVDAEELVPOQGFCDPAPAGGVMVHHRSSSTRSGGDLTLGLEPSEEAAPRSLA 1070

Db 1001 VDAEELVLPYKRI-----NRQGS-----EPCI 1023

QY 1071 PSEGAGSDVFDGLGMAKGLQSLTHDPSPLOQRYSEDPTV-PLPSETDGYVAPLTCSP 1129

Db 1024 PPTGH-----PVRENSITLRNISDPTQNALEKLDLGH----- 1055

QY 1130 QPEYVNPQVRPOP-----PSPRE-----GPLP-AARPAGATLERAKTLPSPKNGVWK 1176

Db 1056 --EYVNPQSETSRRLSDIYNPNVEDLTDGWPVLSLSSQAEATNFSRPEYLVNTNQSL-- 1111

QY 1177 DVFAFGAVENPELYTPQGAAPQPHPPAFSFAFNLYTWDQDPERGAPPTFKGTPT 1236

Db 1112 -PLVSGSGMDPDY---QAG-----YQAAF-----LPQTGALTGNGMFLPA 1148

QY 1237 AENPEYL 1244

Db 1149 AENLEYL 1156

RESULT 9

A36223

kinase-related transforming protein (erbb3) (EC 2.7.1.-) precursor - human

C:Species: Homo sapiens (man)

C>Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000

C:Accession: A36223; 159164

R:Kraus, M.H.; Issing, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.

Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989

A:Title: Isolation and characterization of ERBB3, a third member of the ERBB/epidermal g

A:Reference number: A36223; MUID:90083234; PMID:2687875

A:Accession: A36223

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1342 <KRA>

A:Cross-references: GB:M29366

R:Plowman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todaro, G.J

Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990

A:Title: Molecular cloning and expression of another epidermal growth factor receptor-rel

A:Reference number: 159164; MUID:90311312; PMID:2164210

A:Accession: 159164

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-559, 'G', 561-957, 'F', 959-1063, 'G', 1065-1342 <RES>

A:Cross-references: GB:M34309; NID:gl83990; PIDN:AAA35979.1; PID:g306841

C:Genetics:

A:Gene: GDB:ERBB3; HER3

A:Cross-references: GDB:119880; OMIM:190151

A:Map position: 12q13-12q13

C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog

C:Keywords: ATP; phosphotransferase

F:707-972/Domain: protein kinase homology <KIN>

F:715-723/Region: protein kinase ATP-binding motif

Query Match 34.0%; Score 2305.5; DB 2; Length 1342;

Best Local Similarity 39.5%; Pred. No. 3.9e-92;

Matches 519; Conservative 191; Mismatches 463; Indels 141; Gaps 35;

QY 10 GLILLALLPPCAA--STQVCTGTDMKRLRSPASPTHLDMLRLHYQGCQVVGQNLLEYLPT 67

Db 11 GLILFSLARGSEVGNQSAVCPGTLNGLSVTGAENQYQTLYKLYERCEVGMNLEIVLTGH 70

QY 68 NASLSPLQDIQVQGVVLIHQNQVQPLQRLIRVGTQQLFEDNYALAVLDNGDPLNNTT 127

Db 71 NADLSFLOWIREVTGYVLVAMNEFSLPLNLRVVRGTQVYDGKFAIFVM-----LYNNT 125

QY 128 PVTGASPGGLRELOLRSLTEILKGGVLIQORNPOLCYQDTILMKDIFHKKNQLALTLDTN 187

Db 126 ----NSSHALRQLRLTQLTEILSGGVYIEKNOKLCHMDTIDWRDIVRDRD---AEIVVKD 178

QY 188 RSRACHPCSPMKGRSGWGESSEDCQSLTRITVCAGCG-ARCKGPLPTDCCHEQCAAGCTG 246

Db 179 NGRSCPPCHEVCKG-RCWGPGSEDCQTLTKTICAPQCNHCGFGPNQCHDCACGCSG 237

QY 247 PKHSDCLACHHNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACPYNYLST 306

Db 238 PQDTCDFACRHFENDSGACVPRCPQPLVYNKLTFOLEPNPHTKYQYGGVCVASCAPHNFV-V 296

QY 307 DVGSCSTLVCPHLNQEVTAEDGFORCEKCKSPCARVCYGLGMQYIKANSKF--IGITELE- 363

Db 297 DOTSCVRACPPPKMEVD-KNGLKMCPEPCGGLCPKACEGTG-----SGSRPQTVDSSNIDG 350

QY 364 FAGCKKIFGSLAFELPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPDLS 423

Db 351 FVNCTKILGNLDFLITGLNGDPWHKIPALDPEKLVFRTVREITGYLNTQSPPHMHNFS 410

QY 424 VFQNLQVIRGILHNGAYS-LTQGLGISWGLRLSRELGSGLALIHNNHLCFVHTVPM 482

Db 411 VFSNLTIGRSYLRNRFSLIMKLNVTSLGFRSLKEISAGRIYISANRQLCVHHSLNW 470
QY 483 DQLFRNPHQALLHTA-NRPEDECVGEGLACHOLCARGHGWGPGTQCVCNCSQFLRGECV 541
Db 471 TKVLRGTEERLIDIKHNRPRDCVAEKVCDPLCSCGCGWGPQGCLSCRNVSRGGVCV 530
QY 542 BECRVLQGLPREYVYVNRHCLPCHPECPQNGSVTFCFGEADQCACAHYKDPDFCARCP 601
Db 531 THCNFLNGEPREPAHEAECSCHPECPQMGEGTATCNGSGSDTCAQCAHFRDGHPCVSSCP 590
QY 602 SGVKPDLSPYMIKFPDEBAGCQCPINCHTHSCVDLDDKCPABQRA-----SPLTSIVSA 657
Db 591 HGVLG--AKGPIYKYPDVQNECRCHENCCTGCKGPELQCLGTLLVLIGKTHLTWALT 648
QY 658 VVGILLVVLGVFGILLIKRQOKIR-KYTMRRLLQETELVEPLTPSGAMPNQAEMLK 716
Db 649 IAG--LVVIFMVLGGTFLYWRGRIRQNRAMRYRLERGESIEPLDPS-EKANKVLARIFK 705
QY 717 ETELKVKVLGSGAFQYIK-----ANSKFIGITELVLRNTSPKANKELDEAYVMAG 770
Db 706 ETELKVLKVLGSGVFGTVHKGWIPGESIKIPVCIKVIEDKSGRQSFQAVTDHMLAIGS 765
QY 771 VGSPIYSRLIGICLTSTVQLTQMPYGCLLDHYRENRGRGLSGDILLNWCWQIAKMSYL 830
Db 766 LDHAHIVRLGLCLCGSSQLVLTQYPLGLSLDHYRQHRGALGQLLLNMGVQIAKGMYYL 825
QY 831 EDVRLVHRLDAARNVLKSPNHVKITDFGLARLLDIDETVHADGGKVPKMMALESILR 890
Db 826 BEHGWHRNLARNVLKSPQOVADFVADLLPPDDKQLLYSEAKTPIKMALESIH 885
QY 891 RRFTHQSDVMSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIMV 950
Db 886 GKTHQSDVMSYGVYVWELMTFGAEPYAGRLAEVDPDLLEKGERLAQPICTIDVYIMV 945
QY 951 KCMWIDSECPREFELVSESRWARDQRFVWQNEDLGPA---SPLDSTFYRSLLEDD 1007
Db 946 KCMWIDENIRPTEKELANEFTRWARDPPRYLVIKRES-GPGIAPGPFHGLTNKLEVE 1004
QY 1008 MGLDVAEEYLVPOQGFCDPAPGAGWVHRRSSSTRSGGDLTLGLEP-SEBEAPR 1066
Db 1005 LEPELDLLEDAED-----NLATTLGSLALSPVGTLLNRRPGSQ 1045
QY 1067 SPLAPSEAGSDVFDGLMGAAKGLQSLPTH-DPSPLQYSEDPVPLP-----SETD 1119
Db 1046 SLLSPSSGY-PMNQGNLGCQESAVSGSSRCPRPVSLH-----PMRPGCLASESE 1098
QY 1120 GYVA-----PLTCSPOPE-----YVNPQDVRPQPPSPREP----- 1150
Db 1099 GHVTGSEALQEKVSMCRSRSRSPRPGDSAYHSQRHSLTPVTPPLSPGLEEDVNG 1158
QY 1151 --LPAARPAGATLERAKTSLP-GKNGV-----KQVFAFGGAVENPEYLTPOGGAPOP 1201
Db 1159 YVNPDTHLKTPSREGLTSLVGLSSVLGTEEBED-----EYEYNNRRRRSP-P 1209
QY 1202 HPPAPFAPFDNLYYND-----QDPERGAPPSTFKGTPTAENPEYL 1243
Db 1210 HPPRPSLEBELGYEYMDVGSLSASLGSTQCPPLHPVPIINPTAGTTDEDEYEM 1263

RESULT 10
JC4387
epidermal growth factor receptor homolog precursor - rat
N;Alternate names: ErbB3 protein; HER3 protein
C;Species: Rattus norvegicus (Norway rat)
C;Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998
C;Accession: JC4387
R;Hellyer, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Kolland, J.G.
Gene 165, 279-284, 1995
A;Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protein.
A;Reference number: JC4387; MUID:96096535; PMID:8522190
A;Accession: JC4387
A;Molecule type: mRNA
A;Residues: 1-1339 <HEL>

A;Cross-references: GB:U29339; NID:G915389; PID:G915390
A;Experimental source: liver
A;Note: The authors translated the codon AAC for residue 369 as Thr and GTT for residue 370
C;Comment: This protein is a functional heregulin receptor that transduces signals to the cytoplasm and nucleus.
A;Gene: ErbB3
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog
C;Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane protein
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1339/Product: epidermal growth factor homolog #status predicted <TM>
F;640-659/Domain: transmembrane #status predicted <TM>
F;705-970/Domain: protein kinase homolog <KIN>
F;713-721/Region: protein kinase ATP-binding motif
F;939,1051,1156,1194,1196,1219,1257,1259,1273,1286,1325/Binding site: phosphate (Tyr) (cat)
Query Match 32.7%; Score 2218.5; DB 2; Length 1339;
Best Local Similarity 39.8%; Pred. No. 2.1e-88;
Matches 509; Conservative 168; Mismatches 443; Indels 165; Gaps 36;
QY 3 LAALCRWGLLLALLPPGAA---STVCTGTDMKRLPASPETHLDMRLHLYQCCVQVQGN 59
Db 7 LQVLC-----FLLSLARGSEMGNSQAVCPGTINCLSVTGADNQYQTYLYKLYKECEVMGN 62
QY 60 LEITYLPTNASLSFLQDIOEVQGYVLIANQVRQVPLQRLRIVRGTQTFEDNVALAVLDN 119
Db 63 LEIVLTGHNADLSFLQWIREVTAYVLVAMNEFSLVPLNLRVVRGTQVVDGKEAIFVM-- 120
QY 120 GDEPLNNTTPTVGTASPGGLRELQLRLSITELKGGVLTORNPOLCYQDTILWKDIFHKNQL 179
Db 121 ---LNTNT---NSSHALRQLKFTQLTEILSGGVYIEKNDKLCMDTIDMRDVRVR--- 170
QY 180 ALTLIDNRSRACHPCSPMKSGSRWGSESDCQSLTRTVCAGGC-ARCKGPLPTDCHE 238
Db 171 GAEIVVKNANGANCPPCHEVCKG-RCWGPDPDQCLITKTCAPQCNCRGFGPNQCCHD 229
QY 239 QCAAGCTGPKHSDCLACLHFNHSGICELHCPALVYNTDTTFESMPNPEGRYTFGASCVTA 298
Db 230 ECAGCGSGPDQDTCFACRRFNDSGACVPCRPPLVYNKLTFOLEPNPHTKYQYGGVCVAS 289
QY 299 CPYNYLSTDVGSLVCPHLNQEVTAEDGTORCEKSKPCARVCYL--GMQYIKANSKF 356
Db 290 CPNHFV-VQDTTCVRACPDPKMEVD-KHGLKMCPECGGLCPKACEGTSGSRYTQYVDSN 347
QY 357 IGITELEFAGCKKIFGSLAPLPESPDGPDASNTAPLQEQLOVFELEITGLYIISAWP 416
Db 348 ID----GFVNCYKILGNLDFITGLNVDPHKIPALDPEKLNVRVREITGLYINTQSWP 403
QY 417 DSLPDLVSFQNLQVIRGRILHNGAYS-LTQGLGISWGLSLRLSRELGLALIHNTLHC 475
Db 404 PHMHNFVSFNLTTIGRSLYNRGFSLLIMKNLNTVSLGFRSLKEISAGRVYISANQQLC 463
QY 476 FVHTVPWDQLFRNPHQALLHTA-NRPEDECVGEGLACHOLCARGHGWGPGTQCVCNCSQF 534
Db 464 YHSLNWTLLRGLRPSERLIDIKYDRPLGECLAEKGYCDPLCSCGCGWGPQGCLSCRNV 523
QY 535 LRGEQCEVECRVLQGLPREYVYVNRHCLPCHPECPQNGSVTFCFGEADQCACAHYKDPDP 594
Db 524 SREGCVTHCNFLOGEPREFVHEAQCFSCHEPCLPMEGTSTYNGSGSDACARCAHFRDGP 593
QY 595 FCVARCPGSKVQPDLSYMPIKFPDEBAGCQCPINCHTHSC--VDLDDKCPABQRAASPLT 652
Db 584 HCVNSCPHEILG-AKGPYKYPDAQNECRPCHENCCTGCGNPELQDCLQAEVLMKSPH 641
QY 653 SIVSAVVGILLVVLGVFGILLIKRQOKIR-KYTMRRLLQETELVEPLTPSGAMPNQAO 711
Db 642 LVIAVTG--LAVILMLGSGFLYWRGRIRQNRAMRYRLERGESIEPLDPS-EKANKVL 698
QY 712 MRILKETELRKVKVLGSGAFQYIK-----ANSKFIGITELVLRNTSPKANKELDEA 765
Db 699 ARIFKETELRKVLKVLGSGVFGTVHKGWIPGESIKIPVCIKVIEDKSGRQSFQAVTDH 758
QY 766 YVNVAGVSPYSRLIGICLTSTVQLTQMPYGCLLDHYRENRGRGLSGDILLNWCWQIAK 825

Db 759 LAVGSLDHAHIVRLGLCPSSQLQVLYPLGLSLLDHVVQKHRETLPQQLLNWGVQIAK 818
QY 826 GMSYLEDVRLVHRDLAARNVLVKSPPNHVKITDFGLARLLDIDETEXHADGKGVPIKWMAL 885
Db 819 GMYLLEHSMVHRDLARNVLMKSPSQVQVADFGVADLLPPDKQLLHSEAKTPIKWMAL 878
QY 886 ESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDV 945
Db 879 ESIFGKYTHQSDVWSYGVTVWELMTFGAEPYAGRLAEIPDLLEKGERLAAQPOICTIDV 938
QY 946 YIMVWKWIMDSRCRFRRELVESEFARMARDPQRFVVIQNEIDGPASPLDSTFYRSILLED 1005
Db 939 YMWVWKWIMDENIRTFRELANEFTRMARDPPRYLVIKRAS-GPGTP--PAAPSPVLT 995
QY 1006 DDMGDLVDAEYLVPOQGFCCPDPAFCAGGMVHRRSSRSTSGGGDLTLGLPSEBEE--- 1062
Db 996 KEL-----QEAELEPEL-----DLDLLEAELEG 1021
QY 1063 -----EAPRSLPASEG-----AGSDVFDGLGMAAGKLGSLPT 1097
Db 1022 TSLGSALSUPTGTLTRPGSQSLSPSSGYMPMNQSSLGACLDLSAVLGGREQFSRPSIL 1081
QY 1098 HDPSPLQRYSEDFTVLPSPSTDCGV---APL-----TC-----SPOPE-----YVNQP 1137
Db 1082 H-FIPGR-----PASESEGHVTCSEAELEQKSVCKRSRSPRGRGSAVHSQR 1133
QY 1138 DVPQPSPSPREGP-----LPAARPAGATLERAKTLSP-GKNGVV-----KDV 1179
Db 1134 HSLLTPVPLSPGLBEEDGNGVMPDTHLRGASSREGTLSSVGLSSVLGTBEEDD-- 1191
QY 1180 AFGGAVENPEYLTPOGGAAPQHP 1204
Db 1192 -----EEVEYMNKRKRGSP-PRPP 1209

RESULT 11
TVFVLV
N;Contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase
C;Species: avian leukosis virus, ALV
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 11-Jun-1999
C;Accession: B00643; A00643
R;Nilsson, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1985
A;Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A;Reference number: A00643; MUID:85228222; PMID:2988784
A;Accession: B00643
A;Molecule type: mRNA
A;Residues: 1-698 <N1>
A;Cross-references: GB:M10066; GB:M13881; NID:g211749; PIDN:AAA48763.1; PID:g211750
A;Note: in Genbank entry CHKERBBF, release 109.0, the source is designated as Gallus gal
C;Comment: This protein is synthesized as a gag-env-erbB protein.
C;Genetics:
A;Gene: gag-env-erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
A;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific p
F;1-6/Product: gag protein (fragment) #status predicted <ENV>
F;7-59/Product: env protein (fragment) #status predicted <ENV>
F;60-698/Product: protein-tyrosine kinase erbB #status predicted <ERB>
F;194-459/Domain: protein kinase homology <KIN>
F;202-210/Region: protein kinase ATP-binding motif
F;229/Active site: Lys #status predicted

Query Match 24.4%; Score 1650.5; DB 1; Length 698;
Best Local Similarity 50.1%; Pred. No. 3.3e-64;
Matches 360; Conservative 80; Mismatches 143; Indels 135; Gaps 21;

QY 578 GPEADQCVACAHYKDPFPFCVARGCPGVKPLSYMPIWKFPDEGACQPCINCTHSCVDL 637
Db 60 GP--DHCMKCAHPIDGPHCVKACPAVLGENDTL-VVKYADANAVQLCHPNCTRCCKGP 116
QY 638 DDKGCFAORASPLTSIVSAV--GILLVVVLGVVFGILIKRQOKIRKYMRLLOTEL 696

Db 117 GLEGCP---N3KSTPSIAAGVVGGLLCLVVVGLGIGLYLRRR-HIVRKTRLLRLLQEREL 172
QY 697 VEPLTPSGAMPQAOAMRILKETELRKVKVLGSGAFQO-----YIKANSKF---IGITELV 748
Db 173 VEPLTPSGEAPNQAHLRIKETEFKVKVLGSGAFGVVYKGLWIPGEKVKIPVAIKE-- 230
QY 749 LRENTSPKANKEILDAYVMAGVSGPYVSRLIGICLTSTVQLVTQMLPYGCLLDHVRNR 808
Db 231 LREATSPKANKEILDAYVMASVDNPHVCRLLIGICLTSTVQLITQMLPYGCLLDYIREHK 290
QY 809 GRLGSDLLNCWQIAKMSYLEDVRLVHRDLAARNVLVKSPPNHVKITDFGLARLLDIDE 868
Db 291 DNGSYQLNWCQIANGWYLEERLVRDLAARNVLVKTPOHVKITDFGLARLLGCADE 350
QY 869 TEYHADGGKVPFIKWMALLESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDL 928
Db 351 KEYHAEKGVFIKWMALLESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIIPASEISSV 410
QY 929 LEKGERLPQPICTIDVYIMVWKWIMDSRCRFRRELVESEFARMARDPQRFVVIQ-NED 987
Db 411 LEKGERLPQPICTIDVYIMVWKWIMDSRCRFRRELVESEFARMARDPQRFVVIQODER 470
QY 988 LGPASPLDSTFYRSILLEDMDGDLVDAEYLVPOQGFCCPDPAFCAGGMVHRRSSSTR 1047
Db 471 MHLPSPTDSKFRYTLMEEDMEDIVDAEYLVPHQGF-----NSPST- 513
QY 1048 SGGGDLTLGLEPSEBAPRSP-----APSEGAGSDVFDGLGMAAGKLGSLTPTDHPSP 1102
Db 514 -----SPTPLSSLSATSNATNCID-----RNGQGHVREDSP 548
QY 1103 LQRYSEDPVPLPSET--DGYVAPLTCPOPEYVNPQDVRPPOPPSPREGPLPAARPAAT 1160
Db 549 VQRYSSDPTGNFLESIDDDGFL-----PAPEYVQ--LMPKKPS----- 585
QY 1161 LERAKTLSPGKNGVVKDVP-----AFGGAVENPEYLTPOGGAAPQHP 1205
Db 586 -----TAMVQIYNNISLTAISKLPMSRYQNSHSTAVDNPPEYL-----NTNQSP 631
QY 1206 AFSAPADNLYYDQ-----DPPE-----RGAPPSTFKGTPTAENPEYLGLDVP 1248
Db 632 LAKTVFESSPYIQSGNHQINLDNPDYQODFLPNETKPNGLLKVPAENPEYLRVAAP 689

RESULT 12
TVFVLV
protein-tyrosine kinase (3C 2.7.1.112) erbB - avian erythroblastosis virus (strain H)
C;Species: avian erythroblastosis virus
C;Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 11-Jun-1999
C;Accession: A00644; A38032
R;Yamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.
Cell 35, 71-78, 1983
A;Title: The erbB gene of avian erythroblastosis virus is a member of the src gene family
A;Reference number: A00644; MUID:84026539; PMID:6313229
A;Accession: A00644
A;Molecule type: DNA
A;Residues: 1-604 <YAM>
A;Cross-references: GB:K01216; NID:g209676; PIDN:AAA42400.1; PID:g209678
R;Debuire, B.; Henry, C.; Benaissa, M.; Biserte, G.; Claverie, J.M.; Saule, S.; Martin, J.
Science 224, 1456-1459, 1984
A;Title: Sequencing the erbA gene of avian erythroblastosis virus reveals a new type of
A;Reference number: A38022; MUID:84223957; PMID:6328658
A;Accession: A38022
A;Molecule type: DNA
A;Residues: 1-28,'W',30-119,'P',141-145,'V',147-152 <DEB>
A;Cross-references: GB:K02006
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific p
F;130-395/Domain: protein kinase homology <KIN>
F;138-146/Region: protein kinase ATP-binding motif
F;165/Active site: Lys #status predicted

Matches' 330; Conservative 68; Mismatches 132; Indels 98; Gaps 16;

QY 578 GPEADQCVACAHYKDPFPCVACRSGVKPDLSPYMPKFPDEGACQPCPINCTHSCVDL 637
D 1 GP--DHCWKCAHFDGPHCVKACFAGVLENDTL-VKMYADANAVCQLCHPNCCTRGCKGP 57
QY 638 DDKGCPEQRASPLTSIVSAVV-GILLVVVLGVVFGILIKRQOKIRKYTWRRLLQSTEL 696
D 58 GLECCP---NGSKTPSIAAGVVGGLLVVVGIGLYLRRR-HIVKRTLRLLQREL 113
QY 697 VEPLTPSGAMPNOAQMRILKETELRKVKVLGSGAFQVYK-----ANSKFIGITELVLR 750
D 114 VEPLTPSGEAPNOAHILKETEFKKVKVLGFGAFVYKGLWTPGEKVTIPVAIKELR 173
QY 751 ENTSPKANKEILDAYVMAGVSPYVSRLLGICLTSTVQLVQLMPLYGCLLDHVRNRR 810
D 174 EATSPKANKEILDAYVMASVDNPHVCRLLGICLTSTVQLITQLMPLYGCLLDYIREHKN 233
QY 811 LGSODLLNWCQIAKMSYLEDVRLVHRDLAARNVLKSPNHNKVTITDFGLARLLDIDE 870
D 234 IGSQYLLNWCQIAKGMNILEERHLVHRDLAARNVLKTPQDVKITDFGLAKQLGADEKE 293
QY 871 YHADGGKVPKIMWALESIILRRRFTQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLE 930
D 294 YHAGGKVPKIMWALESIILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGPAREIPDLLE 353
QY 931 KGERLPQPICTIDVIMVIMVCMWIDSECRPRRELSEFRSMARDPQRFVVIQ-NBDLG 989
D 354 KGERLPQPICTIDVIMVIMVCMWIDSECRPRRELSEFRSMARDPQRFVVIQ-NBDLG 989
QY 990 PASPLDSTFRSLLEDMDGLVDAEYLVPOQFFPCDPAPGAGMVHHRHRSSTRS 1049
D 414 LPSPTDSKFYRTLMEEDMEDIVDAEYLVPHQGF-----NSPST--- 454
QY 1050 GGDITLGLSEERAPSP-----APSEGAGSDVFDGLGMGAAGLQSLPHDPFLQ 1104
D 455 -----SPTPLSSLSATSNNSATNDRNG-----H----- 481
QY 1105 RYSEDDTVPPLSEPTDGVVAPLTCSPQPEYVQNPQVPPQSPREGPLPAARPAGAT-LER 1163
D 482 -----PVREDGFL-----PAPEYVQ--LMPKPSVAMVQNIYISLTAISK 523
QY 1164 AKTLPQKGVVVDVFAFGAVENPEYL 1191
D 524 LPMDSRYQN-----SHSTAVDNPEYL 544

RESULT 15

QYFE
epidermal growth factor receptor - fruit fly (Drosophila melanogaster)
N;Contains: protein-tyrosine kinase (BC 2.7.1.112) erbB
C;Species: Drosophila melanogaster
C;Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 11-Jun-1999
C;Accession: A00640; A38021
R;Livnen, B.; Glazer, L.; Segal, D.; Schlesinger, J.; Shilo, B.Z.
Cell 40, 599-607, 1985
A;Title: The Drosophila EGF receptor gene homolog: conservation of both hormone binding
A;Reference number: A00640; MUID:85124611; PMID:2982499
A;Accession: A00640
A;Molecule type: DNA
A;Residues: 1-1330 <LIV>
R;Madsworth, S.C.; Vincent III, W.S.; Bilodeau-Wentworth, D.
Nature 314, 178-180, 1985
A;Title: A Drosophila genomic sequence with homology to human epidermal growth factor re
A;Reference number: A38021; MUID:85137938; PMID:2983232
A;Accession: A38021
A;Molecule type: DNA
A;Residues: 'A', 832-866, 'V', 868-943, 'QTPSLVK' <WAD>
A;Cross-references: EMBL:X02293; NID:g9722; PIDN:CAA26157.1; PID:g929565
C;Comment: This sequence is tentative because the introns have not been identified.
C;Genetics:
A;Gene: FlyBase:Egfr

A;Cross-references: FlyBase:FBgn0003731
A;Map position: 2 57F
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
F;1-732/Domain: extracellular #status predicted <EXT>
F;733-764/Domain: transmembrane #status predicted <TM>
F;765-1330/Domain: intracellular #status predicted <INT>
F;808-1072/Domain: protein kinase homology <KIN>
F;816-824/Region: protein kinase ATP-binding motif
F;122-300,324,363,518,686,695,700/Binding site: carbohydrate (Asn) (covalent) #status pre
F;774/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F;843/Active site: Lys #status predicted
F;1181/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status predicted

Query Match 22.5%; Score 1523.5; DB 1; Length 1330;
Best Local Similarity 28.8%; Pred. No. 1.8e-58;
Matches 400; Conservative 178; Mismatches 416; Indels 393; Gaps 43;

QY 80 VQYVLIANNVROVPLORLIRIVRGTOLEF-----EDNYALAVLDNGDPLNNTTPTVGASP 134
D 38 ITNYIVIGLDLPCTLSYRLQIRGRTLSLSVEEEKYALFV-----TY 81
QY 135 GGLRELQRLSLTEILKGGVLIQNPOLCYQDTILWKDIFHNNQLALTLDITNRSRACHP 194
D 82 SKMYTLEIPDLRDLNGQVGFHNNYLNCHMTTIQWSEIVSNGTDAYNYNDFTAPERECPK 141
QY 195 CSPMKGSRGWGSSSDCQSLTRTVPCAGGCA--RCKGPLTDCCHQCQAGCTGPKHSDC 252
D 142 CHESCTHG-CWGBGPKNCQFKSLTCSPOCAGRCYGPKECCHLFCAGCTGTQKDC 200
QY 253 LACLHFNHSGICELHCPALVYNTDFTFESMPNPEGTYTFGASCVTACPVNYLSTDVGSCT 312
D 201 IACKNFDEAVSKEECPPMRKYNPTTYVLETNPEGKYAYGATCVKECP-GHLLRDNGACV 259
QY 313 LVCPLNQEVTAEDGTQCEKSKPCARVCYGLGMQYIKANSKFIGITEL-----EPAG 366
D 260 RSCPKDMKGG-----CVPCNGPCPKTC-----PGVTVLHAGNIIDFRN 300
QY 367 CKKIFGLAFLPESFDG--DPASNTA-----PQEQLOVFPETLEETGLYLVISAWPDS 418
D 301 CTVIDGNIRILDOTFGSGFDVYANYTGMGRYIPLDPERREVSTVKEITGLYLNIEGTHPQ 360
QY 419 LPDLSVFQNLQVIRGILHNGAY-SLTQGLGHSWLSRLSRLSGLALIHHTHLCFV 477
D 361 FRNLSYFRNLETHGRQLMESFAALAIKSSLSYLEMNLKQISSGVVVIQHNRLCYV 420
QY 478 HTVPDQLFRNPHOALLHTANRDEEC-----504
D 421 SNIRWPAIOKEPEQKVMVNNENLRADLCGKFLTILSVQHNIIMHIFAICREKNHLLGSV 480
QY 505 -----504
D 481 QRGRLGSHGWSVPYVQLQELQFQWHLHRLWLVIQVINSITQDKSNEHQLTDACYSPSVPT 540
QY 505 -----VG 506
D 541 SLTIERARVAIQSAGLAMELEQITARSAMRHSKTLPAEGRQVPRVWFLGVCSASARAGIA 600
QY 507 EGLA-----CHQLCARGHCWGPGTQCVCNCSQFLRGQECVEECRVLQGLPREYV---N 556
D 601 EPLAGRAVCRKCHPLCELCTNYGHEQVGSCKTHYKRREQCEEC-----PADHYTDEE 654
QY 557 ARHCLPCHPECQPNQSVTCFGEADQCVACAHYK-----DPPF-----CVARCPG 603
D 655 QRECFORHEC---NG---CTGPGADDCKSRNFKLFDANETGYPVYNSTWNTCTSKCPLE 708
QY 604 VK-PDLSYMPINWPKFDEGACQPCPINCTHSCVDLDDKGCPEAEQASPLTSIVSAVVGL 662
D 709 MRHVNYQYTAIGPY-----CAASPPRSKITTANLD-----VNMIFITGAV 749
QY 663 LVVLGVVFGI-LIKRQOKIRKYT--MRRLLQETELVEPLTPSGAMPNQAQMRILKETE 719
D 750 LVPTICILCVTVICTQKQKAKKETVKMTWALSREDSSEPLRPSNIGANLCKLRIVKDAE 809

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:47:04 ; Search time 10.5911 Seconds
(without alignments)

4891.279 Million cell updates/sec

Title: SEQ4-710-730-12

Perfect score: 6775

Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPAENPEYGLDVPV 1249

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	6602	97.4	1255	1 ERB2 HUMAN	P04626 homo sapien
2	5811	85.8	1257	1 ERB2 RAT	P06494 rattus norv
3	5795.5	85.5	1254	1 ERB2 MESAU	Q60553 mesocricetu
4	3028	44.7	1210	1 EGFR HUMAN	P00533 homo sapien
5	3002	44.3	1210	1 EGFR MOUSE	Q01279 mus musculu
6	2849.5	42.1	1308	1 ERB4 HUMAN	Q15303 homo sapien
7	2830	41.8	1308	1 ERB4 RAT	Q62956 rattus norv
8	2569.5	37.9	1167	1 XMRK_XIPMA	P13388 xiphophorus
9	2314.5	34.2	1342	1 ERB3 HUMAN	P21860 homo sapien
10	2241.5	33.1	1339	1 ERB3 RAT	Q62799 rattus norv
11	1845	27.2	1426	1 EGFR DROME	P04412 drosophila
12	1633.5	24.1	634	1 ERBB ALV	P00534 avian leuko
13	1595	23.5	703	1 EGFR CHICK	P13387 gallus gall
14	1588	23.4	604	1 ERBB AVIER	P00535 avian eryth
15	1520	22.4	540	1 ERBB AVIEU	P11273 avian eryth
16	1224	18.1	1323	1 LTB2 CAEEL	P24348 caenorhabdi
17	1142.5	16.9	245	1 ERB2 MOUSE	P70424 mus musculu
18	703	10.4	1363	1 ILPR BRALA	Q02466 branchiost
19	676.5	10.0	1372	1 INSR_MOUSE	P15208 mus musculu
20	674	9.9	1383	1 INSR RAT	P15127 rattus norv
21	673	9.9	1382	1 INSR HUMAN	P06213 homo sapien
22	659	9.7	1300	1 IRR_MOUSE	Q9wt14 mus musculu
23	650	9.6	1297	1 IRR_HUMAN	P14616 homo sapien
24	644.5	9.5	1300	1 IRR_CAVPO	P14617 cavia porce
25	643	9.5	1477	1 HTK7_HYDAT	Q25197 hydra atten
26	634	9.4	1607	1 MIPR_LYNST	Q25410 lymanaea sta
27	607	9.0	1367	1 IG1R HUMAN	P08059 homo sapien
28	589	8.7	1373	1 IG1R_MOUSE	Q60751 mus musculu
29	586	8.6	987	1 EPB4_HUMAN	P54760 homo sapien
30	585.5	8.6	1370	1 IG1R RAT	P24062 rattus norv
31	583	8.6	1114	1 RET_HUMAN	P07949 homo sapien
32	574.5	8.5	984	1 EPB1_CHICK	Q07494 gallus gall
33	573	8.5	2146	1 INSR_DROME	P09208 drosophila

RESULT 1	ERB2_HUMAN	ID	ERB2_HUMAN	STANDARD;	PRT;	1255 AA.
34	571.5	8.4	984	1	EPB1 RAT	P09759 rattus norv
35	568	8.4	1390	1	INSR_AEDAE	Q93105 aedes aegyp
36	566	8.4	987	1	EPB4 MOUSE	P54761 mus musculu
37	565.5	8.3	984	1	EPB1_HUMAN	P54762 homo sapien
38	559	8.3	902	1	EPBB_XENLA	Q91736 xenopus lae
39	551.5	8.1	985	1	EPBA_XENLA	Q91571 xenopus lae
40	547.5	8.1	977	1	EPA2_MOUSE	Q03145 mus musculu
41	538.5	7.9	976	1	EPA2_HUMAN	P29317 homo sapien
42	535.5	7.9	985	1	EP4B_XENLA	Q91694 xenopus lae
43	535	7.9	998	1	EPB3_HUMAN	P54753 homo sapien
44	534.5	7.9	986	1	EP4A_XENLA	Q91845 xenopus lae
45	534	7.9	806	1	CEK2_CHICK	P18460 gallus gall

ALIGNMENTS

AC P04626;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell surface receptor HER2) (MLN 19).
GN ERBB2 OR HER2 OR NGL OR NEU.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86118663; PubMed=3003577;
RA Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N., Saito T., Toyoshima K.;
RT "Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth factor receptor."
RL Nature 319:230-234(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=86070181; PubMed=2999974;
RA Coussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A., McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J., Francke U., Levinson A., Ullrich A.;
RT "Tyrosine kinase receptor with extensive homology to EGF receptor shares chromosomal location with neu oncogene."
RL Science 230:1132-1139(1985).
RN [3]
RP SEQUENCE OF 737-1031 FROM N.A.
RX MEDLINE=86016729; PubMed=2995967;
RA Semba K., Kamata N., Toyoshima K., Yamamoto T.;
RT "A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epidermal growth factor-receptor gene and is amplified in a human salivary gland adenocarcinoma."
RL Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).
RN [4]
RP VARIANTS VAL-654 AND VAL-655.
RX MEDLINE=93194196; PubMed=8095488;
RA Ehsani A., Low J., Wallace R.B., Wu A.M.;
RT "Characterization of a new allele of the human ERBB2 gene by allele-specific competition hybridization."
RL Genomics 15:426-429(1993).
CC -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX, ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-ALPHA AND AMPHIRGULIN.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein tyrosine phosphate.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.


```

QY 721 RKVKVLGSGAGQYVKA-----NSKFIGITELVIRENTSPKANKKILDEAYVMAGVGS 773
DQ 721 RKVKVLGSGAGTYGKIWIPIGDNVFK-IPVAIKVLIRENTSPKANKKILDEAYVMAGVGS 779
QY 774 PYVSRLGIGCLTSTVQLVTQVLMYPGCLLDHVRENRRGLSGDNLNWCQIAKGMVSYLEDV 833
DQ 780 PYVSRLGIGCLTSTVQLVTQVLMYPGCLLDHVRENRRGLSGDNLNWCQIAKGMVSYLEDV 839
QY 834 RLVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRRF 893
DQ 840 RLVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRRF 899
QY 894 THQSDVMSYGVWVWELMTFGAKPYDGPAREIPDLLEKGERLPPOPICTIDVYMWKWCW 953
DQ 900 THQSDVMSYGVWVWELMTFGAKPYDGPAREIPDLLEKGERLPPOPICTIDVYMWKWCW 959
QY 954 MIDSECRPRRELVSERWARDPQRFVWQNEIDLGASPLDSTFYRSLLEDDMDGLVD 1013
DQ 960 MIDSECRPRRELVSERWARDPQRFVWQNEIDLGASPLDSTFYRSLLEDDMDGLVD 1019
QY 1014 ABEYLVPOQGFCCPDPAAGAGGMVHRRSSSTRSGGGDLTLGLEPSEEEAPRSPAPSE 1073
DQ 1020 ABEYLVPOQGFCCPDPAAGAGGMVHRRSSSTRSGGGDLTLGLEPSEEEAPRSPAPSE 1079
QY 1074 GAGSDVFDGLGMAAGKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEY 1133
DQ 1080 GAGSDVFDGLGMAAGKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEY 1139
QY 1134 VNQPDVVRQPPSPREGPLPAARAGATLERAKTILSPCKNGVVKDVFAGGAVENPEYLTTP 1193
DQ 1140 VNQPDVVRQPPSPREGPLPAARAGATLERAKTILSPCKNGVVKDVFAGGAVENPEYLTTP 1199
QY 1194 QGGAAPQHPPPAFSPAFDNLVYWDQPPRGPAPPTFTKPTAENPEYLGLOVPV 1249
DQ 1200 QGGAAPQHPPPAFSPAFDNLVYWDQPPRGPAPPTFTKPTAENPEYLGLOVPV 1255

RESULT 2
ID -ERB2 RAT STANDARD; PRT; 1257 AA.
AC P06494;
DC 01-JAN-1988 (Rel. 06, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Epidermal growth factor
DE receptor-related protein).
GN ERBB2 OR NEU.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Neuroblastoma;
RX MEDLINE=86118662; PubMed=3945311;
RA Bargmann C.I., Hung M.-C., Weinberg R.A.;
RT "The new oncogene encodes an epidermal growth factor receptor-related
RT protein";
RL Nature 319:226-230(1986).
RN [2]
RP SEQUENCE OF 852-905' FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=91222560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system.";
RL Neuron 6:691-704(1991).
RN [3]
RP STRUCTURE BY NMR OF 650-668.
RX MEDLINE=92155181; PubMed=1346763;
RA Gullick W.J., Bottomley A.C., Lofts F.J., Doak D.G., Mulvey D.,

```

```

RA Newman R., Crompton M.J., Sternberg M.J.E., Campbell I.D.;
RT "Three dimensional structure of the transmembrane region of the proto-
RL oncogenic and oncogenic forms of the neu protein.";
EMBO J. 11:43-48(1992).
CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS.
CC THE CONSTITUTIVELY ACTIVATED ONCOGENIC VARIANT FORMS A HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; X03362; CAA27059.1; ALT_INIT.
DR PIR; A24562; TVRTNU.
DR HSSP; P11362; LFQK.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; YLP motif.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Proto-oncogene; Disease mutation.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 1257 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
FT DOMAIN 22 654 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 655 677 POTENTIAL.
FT DOMAIN 678 1257 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 159 369 CYS-RICH.
FT DOMAIN 473 646 CYS-RICH.
FT DOMAIN 722 989 PROTEIN KINASE.
FT NP_BIND 728 736 ATP (BY SIMILARITY).
FT BINDING 755 755 ATP (BY SIMILARITY).
FT ACT_SITE 847 847 BY SIMILARITY.
FT DISULFID 196 205 BY SIMILARITY.
FT DISULFID 200 233 BY SIMILARITY.
FT DISULFID 221 236 BY SIMILARITY.
FT DISULFID 225 236 BY SIMILARITY.
FT DISULFID 237 245 BY SIMILARITY.
FT DISULFID 241 253 BY SIMILARITY.
FT DISULFID 256 265 BY SIMILARITY.
FT DISULFID 269 296 BY SIMILARITY.
FT DISULFID 300 312 BY SIMILARITY.
FT DISULFID 316 332 BY SIMILARITY.
FT DISULFID 335 339 BY SIMILARITY.
FT DISULFID 513 522 BY SIMILARITY.
FT DISULFID 517 530 BY SIMILARITY.
FT DISULFID 533 542 BY SIMILARITY.
FT DISULFID 546 562 BY SIMILARITY.
FT DISULFID 565 578 BY SIMILARITY.

```

FT DISULFID 569 586 BY SIMILARITY.
 FT DISULFID 589 598 BY SIMILARITY.
 FT DISULFID 625 625 BY SIMILARITY.
 FT DISULFID 628 636 BY SIMILARITY.
 FT DISULFID 632 644 BY SIMILARITY.
 FT MOD RES 1141 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD RES 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 188 188 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 260 260 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 573 573 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 631 631 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).
 SQ SEQUENCE 1257 AA; 138831 MW; 6129264583011402 CRC64;
 Query Match 85.8%; Score 5811; DB 1; Length 1257;
 Best Local Similarity 85.6%; Pred. No. 4.7e-304;
 Matches 1077; Conservative 53; Mismatches 118; Indels 10; Gaps 4;
 QY 1 MELALCRWGLLALLPPGAASCTGCTDMKRLRASPETHLDMLRHLVQGCQVVOGNL 60
 DB 1 MELANCRWGLLALLPPGIAGTQVCTGDMKRLRASPETHLDMLRHLVQGCQVVOGNL 60
 QY 61 ELTYLPTNASLFLQDIQEVQYVLIHANOVRQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
 DB 61 ELTYVPAASLFLQDIQEVQYVLIHANOVRQVPLQRLRIVRGTLQFEDKYALAVLDNR 120
 QY 121 DPLNNTPTVT-GASPGGLRELQRLSLTEILKGGVLIQRLNPOLCYQDTILKWDIPHKNNQL 179
 DB 121 DPQDNVAASPTGRTPEGLRELQRLSLTEILKGGVLIQRLNPOLCYQDMVLWKVDVPRKNNQL 180
 QY 180 ALTLIDNRSRACHPCSPMCKGSCWGESSDCOSLRTVTCAGGCARCKGLPLTDCCHQ 239
 DB 181 APVDIDNRSRACHPCAPACKDNHGWESPDCQLTGTICTSCARCKGLPLTDCCHQ 240
 QY 240 CAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 299
 DB 241 CAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTTC 300
 QY 300 PYNLSTDVSGCTLVCPHNOEVTABDGTQRCCKSPCARVCYGLGMOYIKANSKFIGI 359
 DB 301 PYNLSTEVSGCTLVCPHNOEVTABDGTQRCCKSPCARVCYGLGMEHLRGARITSD 360
 QY 360 TELEFACGKIFGSLAFIPESFGDGRASNTAPLOEQVPELTETGLVYISAWPDSL 419
 DB 361 NVQFDFGCKIFGSLAFIPESFGDGRSGIAPLPEQVPELTETGLVYISAWPDSL 420
 QY 420 PDLVSFQNLQVIRGILHNGAYSITLQGLGTSWLGRLSLRELGLALIHNTHLCFVHT 479
 DB 421 RDLVSFQNLRIIRGILHDGAYSITLQGLGHSGLRLSLRELGLALIHNAHLCFVHT 480
 QY 480 VPWDLFRNPQALLHTANRPEDE-CVGEGLACHOLCARHCWGPGTQCVCNCSQFLRGQ 538
 DB 481 VPWDLFRNPQALLHNSRPEEDLCVSSGLVNSLCAHGHGCGPGTQCVCNCSHFRLRGQ 540
 QY 539 ECVEECRVLCGLPREYNARHCLCHPECQONSGVTCFGEADQCVACAHYKDPPECVA 598
 DB 541 ECVEECRVWGLPREYVSDKFLCHPECQONSGVTCFGEADQCAACAHYKDSSCVA 600
 QY 599 RCPGSGVPLDSYMPIWFPDEGACQPCINCHTSCVDLDKGCPCAPQASPLTISVAV 658
 DB 601 RCPGSGVPLDSYMPIWKYPDEGICQPCINCHTSCVDLDKGCPCAPQASPTFIATV 660
 QY 659 VGILLVVVLGVGILIKRQOKIRKYTMRLLOETELVEPLTSPGAMPQAOQRILKET 718
 DB 661 VGILLVFLVVGVGILIKRQOKIRKYTMRLLOETELVEPLTSPGAMPQAOQRILKET 720
 QY 719 ELRKVKVLGSAFGQYTKA-----NSKFTGTELVELRENTSPKANKEILDEYVNAV 771
 DB 721 ELRKVKVLGSAFGTVYGIWIPDGENVK-IPVAIKVLENTSPKANKEILDEYVNAV 779
 QY 772 GSPYVSRLLGICLTSTVQLVTPYGLCLDHRVRENRGLSGQDLLNWCMIAGMSYLE 831

DB 780 GSPYVSRLLGICLTSTVQLVTPYGLCLDHRVRENRGLSGQDLLNWCMIAGMSYLE 839
 QY 832 DVELVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGGKVPKIMMALESILRR 891
 DB 840 DVELVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGGKVPKIMMALESILRR 899
 QY 892 RFTHQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLEKGERLPPOPICTIDVYIMVK 951
 DB 900 RFTHQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLEKGERLPPOPICTIDVYIMVK 959
 QY 952 CWMIDSECRPRRELVSERWARDQRFVVIQNEIDLPASPLDSTFYSLLEDMDGDL 1011
 DB 960 CWMIDSECRPRRELVSERWARDQRFVVIQNEIDLPSSPMDSTFYSLLEDMDGDL 1019
 QY 1012 VDAEEYLVPOQGFPCPDPAAGWVHRRHSSTSGGDLTLGLEPSEBAPRSLAP 1071
 DB 1020 VDAEEYLVPOQGFSPDPTPGTSTARRHSSTSGGDLTLGLEPSEBAPRSLAP 1079
 QY 1072 SEGAGSDVFDGDLGMAAKGLQSLPHTDPSPLQRYSEDPTVLPSETDGYVAPLTCSPQ 1131
 DB 1080 SEGAGSDVFDGDLAMGVTKGLQSLSPHLSPLQRYSEDPTLPLPETDGYVAPLACSPQ 1139
 QY 1132 EYVNPQDVPAPQSPREGPLPAARPAAGATLERAKTILSPKNGVVKOVAFPGAVENPEYL 1191
 DB 1140 EYVNPQSEVQPPPLTPEGPLPVRPAGATLERPKTILSPKNGVVKOVAFPGAVENPEYL 1199
 QY 1192 TPQGGAPQPPHAPFAPDNLYYWDQPPPERGAPSTFKGTPTAENPEYLGUDVVP 1249
 DB 1200 VPREGTASPPHSPAFSPADNLYYWDQNSSEOGPPSPNPEGTPTAENPEYLGUDVVP 1257
 RESULT 3
 ERB2_MESAU
 ID ERB2_MESAU STANDARD; PRT; 1254 AA.
 AC Q60553;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor protein-tyrosine kinase erbB-2 precursor (BC 2.7.1.112)
 DE (P188rB2) (NEU proto-oncogene) (C-erbB-2).
 GN ERB2 OR NEU.
 OS Mesocricetus auratus (Golden hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Mesocricetus.
 OX NCBI_TaxID=10036;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Nerve;
 RX MEDLINE=94153007; PubMed=7908275;
 RA Nakamura T., Ushijima T., Ishizaka Y., Nagao M., Arai M.,
 RA Yamazaki Y., Ishikawa T.;
 RT "Cloning and activation of the Syrian hamster neu proto-oncogene.";
 RL Gene 140:251-255 (1994).
 CC - FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX.
 CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
 CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
 CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).
 CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC - SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
 CC (POTENTIAL).
 CC - SUBCELLULAR LOCATION: Type I membrane protein.
 CC - PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
 CC RESIDUES.
 CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

CC EMBL; D16295; BAA03801.1; -
CC HSP; p11362; 1FGK.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; YLP motif.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
DR PROSITE; PS00111; PROTEIN KINASE_DOM; 1.
DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transmembrane; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Proto-oncogene; Disease mutation.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 1254 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
FT DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 653 675 POTENTIAL.
FT DOMAIN 676 1254 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 158 368 CYS-RICH.
FT DOMAIN 472 644 CYS-RICH.
FT DOMAIN 720 987 PROTEIN KINASE.
FT NP_BIND 726 734 ATP (BY SIMILARITY).
FT BINDING 753 753 ATP (BY SIMILARITY).
FT ACT_SITE 845 845 BY SIMILARITY.
FT DISULFID 195 204 BY SIMILARITY.
FT DISULFID 199 212 BY SIMILARITY.
FT DISULFID 236 244 BY SIMILARITY.
FT DISULFID 240 252 BY SIMILARITY.
FT DISULFID 255 264 BY SIMILARITY.
FT DISULFID 268 295 BY SIMILARITY.
FT DISULFID 299 311 BY SIMILARITY.
FT DISULFID 315 331 BY SIMILARITY.
FT DISULFID 334 338 BY SIMILARITY.
FT DISULFID 511 520 BY SIMILARITY.
FT DISULFID 515 528 BY SIMILARITY.
FT DISULFID 531 540 BY SIMILARITY.
FT DISULFID 544 560 BY SIMILARITY.
FT DISULFID 563 576 BY SIMILARITY.
FT DISULFID 567 584 BY SIMILARITY.
FT DISULFID 587 596 BY SIMILARITY.
FT DISULFID 600 623 BY SIMILARITY.
FT DISULFID 626 634 BY SIMILARITY.
FT DISULFID 630 642 BY SIMILARITY.
FT MOD_RES 1139 1139 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1247 1247 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 125 125 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 187 187 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 259 259 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 530 530 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 571 571 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 629 629 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARIANT 658 658 V -> E (IN ONCOGENIC NEU).
FT VARIANT 659 659 V -> E (IN ONCOGENIC NEU).
SQ SEQUENCE 1254 AA; 138252 MW; 974C3791C21F2BE1 CRC64;

Query Match 85.5%; Score 5795.5; DB 1; Length 1254;
Best Local Similarity 85.2%; Pred. No. 3.2e-303;
Matches 1070; Conservative 62; Mismatches 115; Indels 9; Gaps 3;

QY 1 MELAALCRWGLLALLPPGAASQVCTGTDMKRLPASPETHLDMLRHLYQGCVVQGNL 60
DB 1 MELAAWCGWGLLALLSPGASQVCTGTDMKRLPASPETHLDIVRHLYQGCVVQGNL 60

```

```

QY 61 ELTYLPTNASLSFLQDIOEVQGYVLIHNRQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
DB 61 ELTYLPTANATLSFLQDIOEVQGYVLIHNRQVPLQRLRIVRGTLQFEDKYALAVLDNR 120
QY 121 DPLNNTTPTVGASPGGLREQLRLSLTEILKGGVLIQORNPOLCYQDTILKWDIFHNKQLA 180
DB 121 DPLDNVTTATGTPTGELREQLRLSLTEILKGGVLIIRGNPOLCYQDTVLKWDVFRKNQLA 180
QY 181 LTLIDNRSRACHPCSPCKGSRWCSESSDCOSLRTVTCAGGCARCKGLPDTCCHEQC 240
DB 181 PVDIDNRSRACHPCAPACKDNHCWASPEDCOTLTGTIAPRAVPAARLARLPTDCCHEQC 240
QY 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMENPEGRYTFGASCVTACP 300
DB 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMENPEGRYTFGASCVTCP 300
QY 301 YNLSLTDVGSCTLVCPHNOEVTAEADGTORCEKSCPCARVCYGLGMOYKANSKFIGIT 360
DB 301 YNLSLTDVGSCTLVCPHNOEVTAEADGTORCEKSCPCARVCYGLGMOYKANSKFIGIT 360
QY 361 ELEFACKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFETLEETIGVLIYISAWPDSL 420
DB 361 IQEFACKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFETLEETIGVLIYISAWPDSL 420
QY 421 DLSVFQNLQVIRGRIILHNGAYSILTLQGLGSIWGLSLRSLRELGSGLALIHHTHLCFVHTV 480
DB 421 DLSVFQNLQVIRGRIILHNGAYSILTLQGLGSIWGLSLRSLRELGSGLALIHHTHLCFVHTV 480
QY 481 PMDLFRNPHQALLHTANRPEBECVGEGLACHOLCARGHCWGPGTQCVCNCSOFLRGQEC 540
DB 481 PMDLFRNPHQALLHSGNPSEBECGLKDFACYPLCAHGHGWCWPGTQCVCNCSHFLRGQEC 540
QY 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFCVCAR 600
DB 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFCVCAR 600
QY 601 PSGVKPDLSPMPTWKPDEEGACQPCINCTHSCVDLDDKCPAEQASPLTISVAVVG 660
DB 601 PSGVKPDLSPMPTWKPDEEGACQPCINCTHSCVDLDDKCPAEQASPLTISVAVVG 660
QY 661 ILLVVVLGVVFGILIKRQOKIRKYTMRLLOSTELVEPLTPSGAMPNQAQMRILKETEL 720
DB 661 ILLFLVIGVVVFGILIKRQOKIRKYTMRLLOSTELVEPLTPSGAMPNQAQMRILKETEL 720
QY 721 RKVKVLGSGARGQYIKA-----NSKFTGITELVRENTSPKANKEILDEAVVMAGVGS 773
DB 721 RKVKVLGSGARGQYIKAI-----IPVAIKVIRENTSPKANKEILDEAVVMAGLGS 779
QY 774 PYVSRLLLGLCTSTVOLVTQLMPYGCLLDHRVNRGRGLSQDILLNWCMIAGMSYLEDV 833
DB 774 PYVSRLLLGLCTSTVOLVTQLMPYGCLLDHRVNRGRGLSQDILLNWCMIAGMSYLEDV 833
QY 834 RLVRDLAARNVLKSPNHNKVTDFGLARLLDIDETEHADGGKVPFKWMALESILRRRF 893
DB 840 RLVRDLAARNVLKSPNHNKVTDFGLARLLDIDETEHADGGKVPFKWMALESILRRRF 899
QY 894 THQSDVMSYGVVWELMTGAKPYDGI PAREIPDLLEKGERLPPOPICITIDVYMIWKWC 953
DB 900 THQSDVMSYGVVWELMTGAKPYDGI PAREIPDLLEKGERLPPOPICITIDVYMIWKWC 959
QY 954 MIDSECRPRFRLVSEFSRWARDPQRFVVIQNEIDLGSPASPLDSTFYRSILLEDDMDGLVD 1013
DB 960 MIDSECRPRFRLVSEFSRWARDPQRFVVIQNEIDLGSPASPLDSTFYRSILLEDDMDGLVD 1019
QY 1014 ABEYLVPQGGFFCPDPAPGAGGMVHRRHSSSTRSGGDLTLGLEPSEBEAPRSLAPSE 1073
DB 1020 ABEYLVPQGGFFCPDPAPGAGGMVHRRHSSSTRSGGDLTLGLEPSEBEAPRSLAPSE 1079
QY 1074 GAGSDVFDGDLGMAAGKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEY 1133
DB 1080 GAGSDVFDGDLGMAAGKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEY 1139

```


Qy 1134 VNQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVYKDVFFGGAIVENPEYLTP 1193
 Db 1140 VNQEVPRQPPLLTEGPLPPVRPAAGATLERPKTLSPGKNGVYKDVFFGGAIVENPEYLVP 1199
 Qy 1194 QGGAAPQHPPPAPSPAFDNLVYWDQPPERGAPSPSTFKGTPPTAENPEYLGLDVVP 1249
 Db 1200 RGGASAPHP-PPALCPAFDNLVYWDQPPSERGSPPTFGTPTAENPEYLGLDVVP 1254

RESULT 4
 EGFR HUMAN
 ID EGFR HUMAN STANDARD; PRT; 1210 AA.
 AC P00533; P06268; Q14225; Q9UMD7; Q9UMD8; Q9UMG5; Q92795; Q00732;
 AC Q00688; Q9B2S2; Q9B2C9; Q9G2X1; Q9H3C9;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-NOV-2002 (Rel. 41, Last annotation update)
 DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor
 DE protein-tyrosine kinase ErbB-1).
 GN EGFR OR ERBB1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=94219729; PubMed=6328312;
 RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
 RA Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J.,
 RA Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;
 RT "Human epidermal growth factor receptor cDNA sequence and aberrant
 RT expression of the amplified gene in A431 epidermoid carcinoma cells.";
 RL Nature 309:418-425 (1984).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX TISSUE=Placenta;
 RC MEDLINE=95382957; PubMed=7654368;
 RA Ilekis J.V., Stark B.C., Scoccia B.;
 RT "Possible role of variant RNA transcripts in the regulation of
 RT epidermal growth factor receptor expression in human placenta.";
 RL Mol. Reprod. Dev. 41:149-156 (1995).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX TISSUE=Placenta;
 RC MEDLINE=97078686; PubMed=8918811;
 RA Reiter J.L., Mahle N.J.;
 RT "A 1.8 kb alternative transcript from the human epidermal growth
 RT factor receptor gene encodes a truncated form of the receptor.";
 RL Nucleic Acids Res. 24:4050-4056 (1996).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX TISSUE=Placenta;
 RC MEDLINE=97256547; PubMed=9103388;
 RA Ilekis J.V., Gariti J., Niederberger C., Scoccia B.;
 RT "Expression of a truncated epidermal growth factor receptor-like
 RT protein (TEGFR) in ovarian cancer.";
 RL Gynecol. Oncol. 65:36-41 (1997).
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).
 RX TISSUE=Placenta;
 RC MEDLINE=21100872; PubMed=11161793;
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
 RA Schehl Sinclair C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
 RA Balasubramaniam S., Crossley T.R., Magnuson T.R., James C.D.,
 RA Mahle N.J.;
 RT "Comparative genomic sequence analysis and isolation of human and
 RT mouse alternative EGFR transcripts encoding truncated receptor
 RT isoforms.";
 RL Genomics 71:1-20 (2001).
 RN [6]
 RP SEQUENCE OF 575-687 FROM N.A.
 RA Reiter J.L., Threadgill D.W., Danielsen A.J., Schehl C.M.,
 RA Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,

RA Mahle N.J.;
 RT "Human and mouse alternative EGFR transcripts encoding only the
 RT extracellular domain of the receptor.";
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP SEQUENCE OF 713-924 FROM N.A.
 RX MEDLINE=84196372; PubMed=6326261;
 RA Lin C.R., Chen W.S., Krueger W., Stolarsky L.S., Weber W.,
 RA Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;
 RT "Expression cloning of human EGF receptor complementary DNA: gene
 RT amplification and three related messenger RNA products in A431
 RT cells.";
 RL Science 224:843-848 (1984).
 RN [8]
 RP SEQUENCE OF 150-962 FROM N.A.
 RX MEDLINE=84245835; PubMed=6330563;
 RA Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,
 RA Roe B.A., Merlino G.T., Pastan I.;
 RT "Human epidermal growth factor receptor cDNA is homologous to a
 RT variety of RNAs overproduced in A431 carcinoma cells.";
 RL Nature 309:806-810 (1984).
 RN [9]
 RP SEQUENCE OF 1028-1210 FROM N.A.
 RX MEDLINE=85046483; PubMed=6093780;
 RA Simmen F.A., Gope M.L., Schulz T.Z., Wright D.A., Carpenter G.,
 RA O'Malley B.W.;
 RT "Isolation of an evolutionarily conserved epidermal growth factor
 RT receptor cDNA from human A431 carcinoma cells.";
 RL Biochem. Biophys. Res. Commun. 124:125-132 (1984).
 RN [10]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=88217333; PubMed=3329716;
 RA Haley J.D., Whittle N., Bennett P., Kinchington D., Ullrich A.,
 RA Waterfield M.D.;
 RT "The human EGF receptor gene: structure of the 110 kb locus and
 RT identification of sequences regulating its transcription.";
 RL Oncogene Res. 1:375-396 (1987).
 RN [11]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=91107677; PubMed=1988448;
 RA Haley J.D., Waterfield M.D.;
 RT "Contributory effects of de novo transcription and premature
 RT transcript termination in the regulation of human epidermal growth
 RT factor receptor proto-oncogene RNA synthesis.";
 RL J. Biol. Chem. 266:1746-1753 (1991).
 RN [12]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=85270438; PubMed=2991899;
 RA Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;
 RT "Characterization and sequence of the promoter region of the human
 RT epidermal growth factor receptor gene.";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924 (1985).
 RN [13]
 RP SEQUENCE OF 540.
 RA Kohda D.;
 RT Submitted (SEP-1997) to the SWISS-PROT data bank.
 RN [14]
 RP RECEPTOR ACTIVITY.
 RX MEDLINE=84191554; PubMed=6325948;
 RA Mroczkowski B., Mosig G., Cohen S.;
 RT "ATP-stimulated interaction between epidermal growth factor receptor
 RT and supercoiled DNA.";
 RL Nature 309:270-273 (1984).
 RN [15]
 RP PHOSPHORYLATION.
 RX MEDLINE=89278137; PubMed=2543678;
 RA Margolis B.L., Lax I., Kris R., Dombalagian M., Honegger A.M.,
 RA Hawk R., Givol D., Ullrich A., Schlessinger J.;
 RT "All autophosphorylation sites of epidermal growth factor (EGF)
 RT receptor and HER2/neu are located in their carboxyl-terminal tails.
 RT Identification of a novel site in EGF receptor.";
 RL J. Biol. Chem. 264:10667-10671 (1989).
 RN [16]

RP CARBOHYDRATE-LINKAGE SITES ASN-128; ASN-175; ASN-413; ASN-444 AND
RP ASN-528.
RX MEDLINE=96398132; PubMed=8962717;
RA Smith K.D., Davies M.J., Bailey D., Renouf D.V., Hounsell E.F.;
RT "Analysis of the glycosylation patterns of the extracellular domain of
RT the epidermal growth factor receptor expressed in Chinese hamster
RT ovary fibroblasts";
RL Growth Factors 13:121-132 (1996).
RN [17]
RP CARBOHYDRATE-LINKAGE SITES ASN-56; ASN-352; ASN-361; ASN-568 AND
RP ASN-603.
RX MEDLINE=20198209; PubMed=10731668;
RA Sato C., Kim J.-H., Abe Y., Saito K., Yokoyama S., Kohda D.;
RT "Characterization of the N-oligosaccharides attached to the atypical
RT Asn-X-Cys sequence of recombinant human epidermal growth factor
RT receptor";
RL J. Biochem. 127:65-72 (2000).
RN [18]
RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.
RX MEDLINE=98225196; PubMed=9556602;
RA Abe Y., Odaka M., Inagaki F., Lax I., Schlessinger J., Kohda D.;
RT "Disulfide bond structure of human epidermal growth factor receptor";
RL J. Biol. Chem. 273:11150-11157 (1998).
RN [19]
RP REVIEW.
RX MEDLINE=87297456; PubMed=3039909;
RA Carpenter G.;
RT "Receptors for epidermal growth factor and other polypeptide
RT mitogens";
RL Annu. Rev. Biochem. 56:881-914 (1987).
CC -!- FUNCTION: Receptor for EGF, but also for other members of the EGF
CC family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding
CC EGF-like growth factor, Gp30 and vaccinia virus growth factor. Is
CC involved in the control of cell growth and differentiation.
CC -!- FUNCTION: Isoform 2/truncated isoform may act as an antagonist.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Isoform 2 is
CC secreted.
CC -!- ALTERNATIVE PRODUCTS: 4 isoforms; 1/p170 (shown here), 2/p60/
CC truncated isoform/TEGFR, 3/p110 and 4; are produced by
CC alternative splicing.
CC -!- TISSUE SPECIFICITY: Expressed in placenta. Isoform 2 is also
CC expressed in ovarian cancers.
CC -!- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC dimerization, internalization of the EGF-receptor complex,
CC induction of the tyrosine kinase activity, stimulation of cell DNA
CC synthesis, and cell proliferation.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X00588; CAA25240.1; -;
DR EMBL; U95089; AAB53063.1; -;
DR EMBL; U48722; AAC50802.1; -;
DR EMBL; U48723; AAC50804.1; -;
DR EMBL; U48724; AAC50796.1; -;
DR EMBL; U48725; AAC50797.1; -;
DR EMBL; U48726; AAC50798.1; -;
Query Match 44.7%; Score 3028; DB 1; Length 1210;
Best Local Similarity 48.7%; Pred. No. 5.6e-155;
Matches 618; Conservative 174; Mismatches 358; Indels 118; Gaps 26;
QY 11 LLLALLPPGAA--STQVCTGTDMLRLPASPETHLMDLRHLHYQQCVVQGNLELYLPTN 68
DB 14 LLAALCPASALEKKVCQGTSSNKLTLQGTGFEDHFLSLQRMFNNECVVLGNLEITYVQRN 73

QY 69 ASLSFLQDIQEVQGVYLLIAHNVQVPLQRLRIVRGTLQFEDNYALAVLDNGSLNNTTP 128
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 126
74 YDLSFLTKIQEVAGYVYLLIALNTVERIPLENLIQIRGMYEYNSYALAVLSNYD----- 126
QY 129 VTCASPGGLRELQRLSLTEILKGGVLIQRPOLCYODTILWKDIFHNQOLALTLIDTNR 188
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 183
127 ---ANKTGLKELPMRNLQELHGAVFNSNPALCNVESIQWRDIVSDSDFLSNNMSDFQNH 183
QY 189 SRACHPCSPMKSGSRGWESSEDCQSLTRTVACAGCA-RCKGPLPTDCHEQCAAGCTGP 247
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 243
184 LGSCKQKDSCPNGSCWGAEENCQKLTIIQAQCSGRGCRKSPSDCCHNQCAAGCTGP 243
QY 248 KHSDCIACLHFNHSGICELHCPALVTYNTDTFSPMPNPEGRYTFGASCVTACPNYLSLD 307
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 303
244 RESDCLVCRKFRDEATCKDTPPLMLYNTTYQMDVNPESGYSFGATCVKCKPRNVYVD 303
QY 308 VGSCTLVCPLNQOEVTAEQTCRCKSKPCARVCYGLNQYIKANSKFGITELE-FAG 366
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 361
304 HGSCVRACGADSYEM-EEDGVKCKCEGCRKVCNGIGIGEFK-DLSLSINATNIKFKN 361
QY 367 CKKIFGSLAFIPESFDGDPASNTAPLOPELOQVETLEETGYLYTSAMPDLSPLDSVFO 426
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 421
362 CTSISGDLHLIPVAFRGDSFTHPPDLPQELDLKTVKEITGELLQAWPENRTDLHAFE 421
QY 427 NLQVIRGRILHNGAYSLTLQGLIGISWGLSLRSLRELSGLALIIHHNTHLCFVHTVPMQDLF 486
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 481
422 NLEIIRGRTKHQGFSLAVVSLNITSGLSLRSLKEISDGDVIIISGNKNCVCYANTINWKKLF 481
QY 487 RNPQALLHTANRPEDECVEGLACHQLCARGHCWGPPTQCVCNCSQFLRGQECVECRV 546
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 541
482 GTSQGTQKIISNGENSKATGVCHALCSPEGCGPEPRDCVSRNVSREGCVDCRNL 541
QY 547 LQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPFCVACRCPGVKP 606
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 601
542 LEGEPREFVENSICIOCHPECLPQANNITCTGRGPNICIOCAHYIDGPHCVKTCPCAGVG 601
QY 607 DLSYMPIWKFPEDEGACQPCFPCINCTHSCVDLDDKGPFAEQRASPLTSIVSAVVG---ILL 663
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 658
602 ENNTL-VWKYADAGHCHLCHPNCTYCTGPGLEGCTNGPKIP--SIATGMVGALLLL 658
QY 664 VVVLGVVFGILIKRQOKIRKYMRLLOSTELVEPLTPSGAMPNQOMRILKETELRVK 723
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 715
659 VVALGIG---LFMRRRHIVKRTLRLLQRELVEPLTPSGEAPNOALLILKETEFKI 715
QY 724 KVLGSGAFGQ----YIKANSKF---IGITELVRENTSPKANKEILDEAYVNAVAGSPY 775
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 773
716 KVLGSGAFGTVYKGLMPEGEKVKIPVAIKE--LREATSPKANKEILDEAYVNAVSDNPH 773
QY 776 VSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRLGSLQDLNWCMIQAKMSYLEVDRL 835
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 833
774 VCRLLGICLTSTVQLITQLMPFGCLLDVYREHKDNIGSQVLLNWCQVQAKGMYLEDRRL 833
QY 836 VHRDLAARNVLKSPNHVKITDFGLARLLDIDTEYHADGGKVPKIKMALESILRRPFT 895
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 893
834 VHRDLAARNVLKTPQHKITDFGLAKLLGABEKEYHAEGKVPKIKMALESILHRTYTH 893
QY 896 QSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITDVIYMLVWKCM 955
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 953
894 QSDVWSYGVTVWELMTFGSKPYDGIIPASELSSLEKGERLPQPPICITDVIYMLVWKCM 953
QY 956 DSECRPRFRELVSFSESRMARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDGMDLVDA 1014
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 1013
954 DADSRPKFRELIIIEFSKMARDPQYLVIOGDERMHLPSPTDSNFYRALMDEEDMDVDVA 1013
QY 1015 EYVLVPOQGFPCDDPAPAGAGVMVHRRHSSTRSGGDLTLGLEPSEEEAPRSPAPSEG 1074
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 1039
1014 DEYLVPOQGF-----SSPSTRPPLLSSLS 1039
QY 1075 AGSDVFDGDLGMAAGKIQSLPHTDPSPLORYSEDTVPPLPSET--DGYVAPLTCPOPE 1132
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 1091
1040 ATSN--NSTVACIDRNLGLOSCPIKEDSFQRYSDPTGALTEDSIDDTFL-----VPVE 1091

Qy 1133 YVQPDVVRQPPSPREGPLPAARPGATLERAKTLSPKNGVWVKDVFAGGAVENPYL- 1191
 Db 1092 YINQ-SVPKRPAQSVQNPVYHQLNP-----AFSRDHYQD--PHSTAVGNPEYL 1140
 Qy 1192 TPQGAAPQHPHPPAFSPAFNLYWQO-----DP-----PERGAPPTFKGTP 1235
 Db 1141 TVQ-----PTCVNSTEDSPAHWAQKQSGHQISLDNPDYQDFFPKPKNGIFKGS- 1190
 Qy 1236 TAENPEYL 1243
 Db 1191 TAENPEYL 1198
 RESULT 5
 EGF_MOUSE
 ID_EGFR_MOUSE STANDARD; PRT; 1210 AA.
 AC Q01279;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Epidermal growth factor receptor precursor (EC 2.7.1.112).
 GN EGFR.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c; TISSUE=Liver;
 RX MEDLINE=93026370; PubMed=1408137;
 RA Avivi A., Skorecki K., Yavon A., Givol D.;
 RT "Promoter region of the murine fibroblast growth factor receptor 2
 (bek/KGFR) gene."
 RL Oncogene 7:1957-1962(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c, and CD-1; TISSUE=Liver, and Decidua;
 RX MEDLINE=93126380; PubMed=7678348;
 RA Paria B.C., Das S.K., Andrews G.K., Dev S.K.;
 RT "Expression of the epidermal growth factor receptor gene is regulated
 in mouse blastocysts during delayed implantation."
 RL Proc. Natl. Acad. Sci. U.S.A. 90:55-59(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c; TISSUE=Liver;
 RA Hibbs M.L.;
 RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=B6/C3; TISSUE=Liver;
 RX MEDLINE=94170986; PubMed=8125255;
 RA Luetsteke N.C., Phillips H.K., Qiu T.H., Copeland N.G., Earp H.S.,
 RA Jenkins N.A., Lee D.C.;
 RT "The mouse waved-2 phenotype results from a point mutation in the EGF
 receptor tyrosine kinase."
 RL Genes Dev. 8:399-413(1994).
 RN [5]
 RP SEQUENCE OF 1-714 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=91232866; PubMed=2030916;
 RA Avivi A., Lax I., Ullrich A., Schlesinger J., Givol D., Morse B.;
 RT "Comparison of EGF receptor sequences as a guide to study the ligand
 binding site."
 RL Oncogene 6:673-676(1991).
 RN [6]
 RP SEQUENCE OF 969-1117 FROM N.A.
 RC STRAIN=C3H;
 RX Avinger D.P., Serrero G.;
 RA Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
 AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
 VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein

CC tyrosine phosphate.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
 CC dimerization, internalization of the EGF-receptor complex,
 CC induction of the tyrosine kinase activity, stimulation of cell DNA
 CC synthesis, and cell proliferation.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X78987; CAA55587.1; -
 CC EMBL; U03425; AAA17899.1; -
 CC EMBL; X59698; CAA42219.1; -
 CC EMBL; L06864; AAA53029.1; -
 CC EMBL; Z12608; CAA78249.1; -
 CC HSSP; P11362; IFGK.
 CC MGD; MGI:95294; Egfr.
 CC InterPro; IPR000494; EGFR_L_domain.
 CC InterPro; IPR000719; Euk_pkinase.
 CC InterPro; IPR002174; Furin-like.
 CC InterPro; IPR001245; Tyr_pkinase.
 CC Pfam; PF00069; pkinase; 1.
 CC Pfam; PF00757; Furin-like; 1.
 CC Pfam; PF01030; Recep_L_domain; 2.
 CC ProDom; PD000001; Euk_pkinase; 1.
 CC SMART; SM00261; FU; 3.
 CC SMART; SM00219; TyrKC; 1.
 CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 CC PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
 CC Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
 CC Tyrosine-protein kinase; ATP-binding; Phosphorylation; Repeat.
 CC SIGNAL 1 24 POTENTIAL
 CC CHAIN 25 1210 EPIDERMAL GROWTH FACTOR RECEPTOR.
 CC DOMAIN 25 647 EXTRACELLULAR (POTENTIAL).
 CC TRANSMEM 648 670 POTENTIAL.
 CC DOMAIN 671 1210 CYTOPLASMIC (POTENTIAL).
 CC REPEAT 75 300 APPROXIMATE.
 CC REPEAT 390 600 APPROXIMATE.
 CC DOMAIN 1028 1071 SER-RICH.
 CC DOMAIN 714 981 PROTEIN KINASE.
 CC NP_BIND 720 728 ATP (BY SIMILARITY).
 CC BINDING 747 747 ATP (BY SIMILARITY).
 CC ACT_SITE 839 839 BY SIMILARITY.
 CC DISULFID 190 199 BY SIMILARITY.
 CC DISULFID 194 207 BY SIMILARITY.
 CC DISULFID 215 223 BY SIMILARITY.
 CC DISULFID 219 231 BY SIMILARITY.
 CC DISULFID 232 240 BY SIMILARITY.
 CC DISULFID 236 248 BY SIMILARITY.
 CC DISULFID 251 260 BY SIMILARITY.
 CC DISULFID 264 291 BY SIMILARITY.
 CC DISULFID 295 307 BY SIMILARITY.
 CC DISULFID 311 326 BY SIMILARITY.
 CC DISULFID 329 333 BY SIMILARITY.
 CC DISULFID 506 515 BY SIMILARITY.
 CC DISULFID 510 523 BY SIMILARITY.
 CC DISULFID 526 535 BY SIMILARITY.
 CC DISULFID 539 555 BY SIMILARITY.
 CC DISULFID 558 571 BY SIMILARITY.
 CC DISULFID 562 579 BY SIMILARITY.
 CC DISULFID 582 591 BY SIMILARITY.
 CC DISULFID 595 617 BY SIMILARITY.
 CC DISULFID 620 628 BY SIMILARITY.
 CC DISULFID 624 636 BY SIMILARITY.
 CC MOD_RES 680 680 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
 CC MOD_RES 1092 1092 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).

FT	MOD_RES	1110	1110	PHOSPHORYLATION (AUTO-)	(BY SIMILARITY)
FT	MOD_RES	1172	1172	PHOSPHORYLATION (AUTO-)	(BY SIMILARITY)
FT	MOD_RES	1197	1197	PHOSPHORYLATION (AUTO-)	(BY SIMILARITY)
FT	CARBOHYD	128	128	N-LINKED (GLCNAC)	(POTENTIAL)
FT	CARBOHYD	175	175	N-LINKED (GLCNAC)	(POTENTIAL)
FT	CARBOHYD	196	196	N-LINKED (GLCNAC)	(POTENTIAL)
FT	CARBOHYD	352	352	N-LINKED (GLCNAC)	(POTENTIAL)
FT	CARBOHYD	413	413	N-LINKED (GLCNAC)	(POTENTIAL)
FT	CARBOHYD	444	444	N-LINKED (GLCNAC)	(POTENTIAL)
FT	CARBOHYD	528	528	N-LINKED (GLCNAC)	(POTENTIAL)
FT	CARBOHYD	568	568	N-LINKED (GLCNAC)	(POTENTIAL)
FT	CARBOHYD	603	603	N-LINKED (GLCNAC)	(POTENTIAL)
FT	CARBOHYD	623	623	N-LINKED (GLCNAC)	(POTENTIAL)
FT	CONFLICT	19	19	C -> S (IN REF. 2)	
FT	CONFLICT	539	539	C -> W (IN REF. 5)	
FT	CONFLICT	991	991	L -> F (IN REF. 4)	
FT	CONFLICT	1116	1117	HP -> DR (IN REF. 6)	
FT	SEQUENCE	1210	AA; 134853 MW; 690E20D46DF2D2F5	CRC64;	
Query Match					44.3%; Score 3002; DB 1; Length 1210;
Best Local Similarity					48.6%; Pred. No. 1.4e-153;
Matches					620; Conservative 168; Mismatches 365; Indels 122; Gaps 28;
QY	11	LLIALLPPGAA--STQVCTGTGDMKRLPASPETHLMDRLHLYQCCVQVQGNLETLPTN	68		
Db	14	LLTALCAAGGALEKKVCQGTNRNLTLQGTGFEDHFLSLQRMNNEVVLGNLEITTYVQRN	73		
QY	69	ASLSFLQDIQVQCVYLIHNOVQVPLQRLRIVRGTQLPEDNVALAVLNDGDPNNTP	128		
Db	74	YDLSFLTKIQVAGVYLIHNTVERIFLENLIQIRGNALYENTYVALILSN-----	124		
QY	129	VTGASPGGLRELQRLSLEILKGVLIRNPQLCYQDTILMKDT-----FKNNNALTLI	184		
Db	125	-YGNRTGLRELPMNLQELILGAVRESNPNILCNMTIQWRDIVQNVFMSNMDL---	180		
QY	185	DTNRSRACHPCSPCKSGRWGESSEDCQSILTRTVVAGGCA-RCKGPLPTDCHEQCAAG	243		
Db	181	-QSHPSPCPKDCPSCPNGSCWGGEENCQKLTIKCAQCQSHRCRGRSPSDCCNQCAAG	239		
QY	244	CTGPKHSDCLACLFHNSGICELHCPALVTYNTDTPESMNPGEGRYTFGASCVTACPVNY	303		
Db	240	CTGPRESDELVCQFQDEATCKDTCPPLMLYNTTYQMDVNPBGSKYFGATCVKCKPRNY	299		
QY	304	LSTDVSGCTLVCPLHNOEVTAEEDTQRCCKSKPCARVCYGLGMQYIKANSKFTIGITELE	363		
Db	300	VWTDHGSCVRACGPDYEV-EEDGIRKCKCDGCRKVCNGIGIGEFK-DTLSINATNIK	357		
QY	364	-FAGCKIIFGLAPLPESFGDPASNTAPLOPEQLQVFTELHEITGYLYISAWPDSLPDL	422		
Db	358	HFKYCTAISGDHLPLVAFKGDSPTRTPPLDPRELEILTKVKEITGFLLIQAWPDNWTDL	417		
QY	423	SVFONLQVIRGRIHLHGAYSITLQGLGISWGLRLSRELGLALIHNTHLCFVHTVPW	482		
Db	418	HAPENLEIRGRTHQHQFSLAVVGLNITSLGRSLKEISDGDVITSGNRLCYANTINW	477		
QY	483	DQLFRNPQALLHTANPEDECVGEGIALACHARGCWGPGTQCVCNCSOFLRGOCVE	542		
Db	478	KKUFGTGNQTKIMNNAEKDKAVNHVCPNLCSEGCWGPEPRDCVQCNVSGRECV	537		
QY	543	ECRVQLGLPREYNARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPPEFCVARCPS	602		
Db	538	KCNILEGEPREFENSECICQHPCELPQAMNITCTGRGPDNCICQAHYIDGPHCVKTCPA	597		
QY	603	GVKPDLSYMPWPKPDEEGACQPCINCHSCVDLDDKGPCEQASPLTSIYSAVGIL	662		
Db	598	GIMGENNLT-LVWKYADANNVCHLCHANCTYGCAGPLQGCVEGWPSGPKFISATIGVGL	656		
QY	663	LVVVLGVVFGI-LIKRQKIRKTYMRLLQETELVEPLTPSGAMPNQAMRILKETELR	721		
Db	657	LFIVV-VALGIGLPMRRHVRKTLRLQLQERLELVEPLTPSGEAPNQAHRLIKETEFK	715		
QY	722	KVKVLGSGAFQ-----YIKANSKF-----IGITELVLRENTSPKANKEILDEAYVMAGVGS	773		

Db	716	KIVLGSAGFETVYKGLWPEGEKXIPVAIKS--LREATSPKANKEILDEAYVMASVDN	773		
QY	774	PVYSRLIGICLTSTVOLVTOLMPYGCLLLDHVRNRRGLSGQDLLNWCMIKAGMSYLEDV	833		
Db	774	PHVCRLLIGICLTSTVOLITOLMPYGCLLDVVRHKNIGSQYLLNWCMIKAGMSYLEDR	833		
QY	834	RLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMWMALESILRRRF	893		
Db	834	RLVHRDLAARNVLKTPQHVKITDFGLAKLLGAEKEHYHAEGGKVPKMWMALESILHRIY	893		
QY	894	THOSDVMSYGVTVWELMTFCAGKPYDGIIPAREIPDLLEKGERLIPQPPCTTIDVYMWKWC	953		
Db	894	THOSDVMSYGVTVWELMTFGSKPYDGIIPASDISSILEKGERLIPQPPCTTIDVYMWKWC	953		
QY	954	MIDSECRPRFRELVSFMRMARDPQRFVVIQ-NEDLGPASPLDSTFYRSLEDDDMGDLV	1012		
Db	954	MIDADSRPFRELILFESKWARDPQRYLVIOGDERMHLPSPTNSFTRALMDEDMEDVV	1013		
QY	1013	DAEYILVPOQGFPCPDPAAGAGMVHRRSSSTRSGGGDLTLGLEPSEEEAEPRLAPS	1072		
Db	1014	DADEYLIPQCGPF-----NSPST-----SRTPLSS	1039		
QY	1073	EGAGSDVFDGLMGAAKGLQSLPTHDPSPLOQRYSEDPTVPLPSET--DGYVAPLTCSPQ	1130		
Db	1040	LSATSN-----NSTVACINRNGSCRVEDAFLOQRYSSDPTGAVTEDNIDDAFL-----	1089		
QY	1131	PEVYNOPDVRFPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAGGAVENPEY	1190		
Db	1090	PEVYNO-SVPRPAGSVQNPVYHNPQHP-----APGRDLHYQN--PHSNAVGNPEY	1138		
QY	1191	L-TPQGAAPQHPHPPAFPAFNLVYWDQ-----DP-----PERGAPPSTFKG	1233		
Db	1139	LNTAQ-----PTCLSSGFNSPALWIKQKSHQMSLONPDYQDFFPKETKPNGIFKG	1189		
QY	1234	TPTAENPEYGLDVP	1248		
Db	1190	-PTAENAEYLRVAPP	1203		
RESULT 6					
ERB4_HUMAN	ID	ERB4_HUMAN	STANDARD;	PRT;	1308 AA.
AC	Q15303;				
DT	15-DEC-1998	(Rel. 37, Created)			
DT	15-DEC-1998	(Rel. 37, Last sequence update)			
DT	15-JUN-2002	(Rel. 41, Last annotation update)			
DE	Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112)				
DE	(p180erbB4) (Tyrosine kinase-type cell surface receptor HER4)				
GN	ERB4 OR HER4				
OS	Homo sapiens (Human)				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A. (ISOFORM JM-A)				
RC	TISSUE=Breast carcinoma;				
RX	MEDLINE=93189574; PubMed=8383326;				
RA	Plowman G.D., Culouscou J.-M., Whitney G.S., Green J.M., Carlton G.W.,				
RA	Foy L., Neubauer M.G., Shoyab M.;				
RT	"Ligand-specific activation of HER4/p180erbB4, a fourth member of the				
RT	epidermal growth factor receptor family."				
RL	Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750(1993).				
RN	[2]				
RP	SEQUENCE FROM N.A. (ISOFORMS JM-A AND JM-B)				
RC	TISSUE=Fetal brain;				
RX	MEDLINE=97476287; PubMed=9334263;				
RA	Elenius K., Corfas G., Paul S., Choi C.J., Rio C., Plowman G.D.,				
RA	Klagabrun M.;				
RT	"A novel juxtamembrane domain isoform of HER4/erbB4. Isoform-specific				
RT	tissue distribution and differential processing in response to				
RT	phorbol ester."				
RL	J. Biol. Chem. 272:26761-26768(1997).				


```
Db 536 ONLYDGBREFRENGSICVECDPQCEKEDDGLTCHGPGPDNCTKCSHPKQDPCNVCEKCPD 595
Qy 603 GVXPDLSYMPKPFDEBACQPCPCINCHTSCVDLDDKGC-----PARQRASPL 651
Db 596 GLOGANSF--IFKYADPDRECHPCPNCTCGCNGPTSHDCIYYPTWGTSTLPQAR--TPL 652
Qy 652 TSIVSAVV-GILLVVVLGVGFIILKROQKIRKYTWRRLLQETELVELPELTPSGAMPNOA 710
Db 653 --IAAGVIGGFILVIVGLTFAVYVRKRSIK-KRALRRFT--ETELVELPELTPSGTAPNOA 708
Qy 711 QMRILKETELRKVKVLGSGAGQVIK-----ANSKFTIGITELVRLRENTSPKANKEILDE 764
Db 709 QLRILKETELRKVKVLGSGAGQVTKYGIWVPEGETVKPIVAIKILNETTGPKANVEFWE 768
Qy 765 AYVMAGVSPXVSRLLGICLTSTVQLVTQLMPYGCGLDHDVRENRGLGSDQLLNCWQIA 824
Db 769 ALIWMASHDHPHLVRLGLVCLSPITQLVTQLMPHGCLEYYVHEKDNIGSQLLNWCQIA 828
Qy 825 KMSYLEVDRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMA 884
Db 829 KGMYLEERRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMA 888
Qy 885 LESILRRFTHOSDVWSVGTWELMTFGAKPYDGIIPARETPDLLEKGERLPQPPICITID 944
Db 889 LECHYKRFTHOSDVWSVGTWELMTFGKPYDGIIPARETPDLLEKGERLPQPPICITID 948
Qy 945 VYIMVVKCWMIDSECRPRFRELVSFSEFMRDPPQFVVIQNEED-LGPASPLDSTFYRSLL 1003
Db 949 VYVMVVKCWMIDADSRRPKELAAEFSEMRDPPQYLVIOGDDRMKLPSPNDSKFFQNL 1008
Qy 1004 EDDMGDLVDAEYLVVQQGFCDPDPAGAGVHHRHSRSTSGGGLTLGLEPSEEE 1063
Db 1009 DEEDLEDMDAEYLVV--QAFNIPP-----IYTSRARIIDSNRS-----EIGHSPPPAY 1056
Qy 1064 APRS-----PLAP-SEGAGSDVFDGDLGMAKGLQS 1094
Db 1057 TPMSEGNFVRDGGFAEAGVSVPYRAPTSITPEAPVAQATAFIDDCSCNGTLRKPEVA 1116
Qy 1095 LPTHDPSPLOYSDDPTVPLPS-----ETDGVAPLTCSPQEPYVNPQDVRPQPPSPR 1147
Db 1117 PHVEDSTQRYSDPTVFAPERSPRGELDEGYWTPMRDKPQEYLNPE-----1167
Qy 1148 EGPLPAARPAGATLERAKTLPSPGNVGVKVFAGGAVENPEYLTPOGGAAPPHPPPA- 1206
Db 1168 ENPFVSR--KNGDLQ-----ALDNPYHNASNG-----PPKAE 1199
Qy 1207 -----FSPAFDNLVYWDQPPPERGA--PSTTF 1231
Db 1200 DEYNEPEYLNFTANTLGKAEYLNKILMSPEKAKAFDNDPDYWNHSLPPRSTLQHPDYL 1259
Qy 1232 KGTP-----AENPEYL 1243
Db 1260 QEYSTKYFYKQNGRIRPIVAENPEYL 1285

RESULT 7
ID_ERB4_RAT STANDARD; PRT; 1308 AA.
AC Q62956; Q922N7;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112).
GN ERBB4 OR TYRO-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=98221155; PubMed=9553078;
RA Zhao Y.-Y., Sawyer D.R., Baliga R.R., Opel D.J., Han X.,
```

```
RA Marchionni M.A., Kelly R.A.;
RT "Neuregulins promote survival
RT Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT ventricular myocytes.";
RL J. Biol. Chem. 273:10261-10269(1998).
RN [2]
RP SEQUENCE OF 848-901 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=91222560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system.";
RL Neuron 6:691-704(1991).
RN [3]
RP SEQUENCE OF 1031-1198 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Fromert P.W., Kim S.S., Corbett J.A.;
RT "Expression of neuregulins and their putative receptors, ErbB2 and
RT ErbB3, is induced during Wallerian degeneration.";
RL J. Neurosci. 17:1642-1659(1997).
CC -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
CC 2. NRG-3. HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
CC NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC RECEPTORS (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC HEART.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF041838; AAD08899.1; -.
CC EMBL; U52531; AAC53051.1; -.
CC HSSP; P11362; IFGK.
CC InterPro; IPR000494; EGFR_L_domain.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR001245; Tyr_pkinase.
CC InterPro; IPR004019; YLP_motif.
CC Pfam; PF00757; Furin-like; 1.
CC Pfam; PF00069; pkinase; 1.
CC Pfam; PF01030; Recep_L_domain; 2.
CC Pfam; PF02757; YLP; 2.
CC PRINTS; PR00109; TYRKINASE.
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00261; FU; 4.
CC SMART; SM00219; TyrKc; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
CC Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 1308 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
FT DOMAIN 26 651 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 652 675 POTENTIAL.
FT DOMAIN 676 1308 CYTOPLASMIC (POTENTIAL).
```

```
FT DOMAIN 186 334 CYS-RICH.
FT DOMAIN 496 633 CYS-RICH.
FT DOMAIN 718 985 PROTEIN KINASE
FT NP_BIND 724 732 ATP (BY SIMILARITY).
FT BINDING 751 751 ATP (BY SIMILARITY).
FT ACT_SITE 843 843 BY SIMILARITY.
FT DISULFID 189 197 BY SIMILARITY.
FT DISULFID 193 205 BY SIMILARITY.
FT DISULFID 213 221 BY SIMILARITY.
FT DISULFID 217 229 BY SIMILARITY.
FT DISULFID 230 238 BY SIMILARITY.
FT DISULFID 234 246 BY SIMILARITY.
FT DISULFID 249 258 BY SIMILARITY.
FT DISULFID 262 289 BY SIMILARITY.
FT DISULFID 293 304 BY SIMILARITY.
FT DISULFID 308 323 BY SIMILARITY.
FT DISULFID 326 330 BY SIMILARITY.
FT DISULFID 503 512 BY SIMILARITY.
FT DISULFID 507 520 BY SIMILARITY.
FT DISULFID 523 532 BY SIMILARITY.
FT DISULFID 536 552 BY SIMILARITY.
FT DISULFID 555 569 BY SIMILARITY.
FT DISULFID 559 577 BY SIMILARITY.
FT DISULFID 580 589 BY SIMILARITY.
FT DISULFID 593 614 BY SIMILARITY.
FT DISULFID 617 625 BY SIMILARITY.
FT DISULFID 621 633 BY SIMILARITY.
FT MOD_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1188 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 138 138 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 174 174 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 253 253 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 358 358 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 410 410 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 473 473 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 495 495 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 548 548 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 620 620 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 1062 1062 S -> N (IN REF. 3).
FT CONFLICT 1080 1082 PYT -> SYR (IN REF. 3).
SQ SEQUENCE 1308 AA; 146957 MW; D944BB0996A08B41 CRC64;

Query Match 41.8%; Score 2830; DB 1; Length 1308;
Best Local Similarity 43.9%; Pred. No. 2.5e-144;
Matches 591; Conservative 191; Mismatches 399; Indels 166; Gaps 29;

Qy 1 MELA-ALCRWGLL--ALLPPGAASVCTCTDMKRLPASPETHLDMLRLHYOGCQVQ 57
Db 1 MKLATGLVWGSLLVAARTVPSASQSVACATENKLSLSLEQQYRALRYENCVEVM 60

Qy 58 GNLELTLYPTNASLFLQDIQEOGYVLIHNOVQVPLQRLRIVRGTLQFEDNYALAVL 117
Db 61 GNLEITSIEHNRDLNFLRSIREVTGYVLVALNQRYLPLENLRIIRGTLKYEDRYALAIF 120

Qy 118 DNGDPLNNTTPTVGTASPGGLRELQRLSLETLKGVLIQRPQLCYQDTILWKDIFHKN 177
Db 121 LNYRKDGNF-----GLOELGKNLTILNGGVYVQNKFLCYADTIHWQDIVRNPW 171

Qy 178 QLALTLIDTNRSRACHPCSPKSGRCSGESSEDCQSILTRVTVACGGC-ARCKGPLPTDCC 236
Db 172 PSNMTLVSTIGSGGGRCHKSCGTG-RCWGPTEHCOITLRTVCAEQDGRCYGYVSDCC 230

Qy 237 HEQCAAGCTGPKHSDCLACLFHNSHIGICELHCPALVTYNTDTFSPMPNPEGRYTFGASCV 296
Db 231 HREAGCGSGPKDTCFACMNFNDSGACVTCQPTQFVNPPTTFQLEHNFNAKYTYGAFCV 290

Qy 297 TACPYNLYSTDVSGCTLVCPHNOEVTAEADGTQRCCKSKPCARVCVGLQMYIKANSKF 356
Db 291 KCCPHNFV-VDSSSCVRCAPSSKMEV-EENGIRKMKCPCTDCPRACDGDIGTGSLSMAQTV 348
```

RESULT 8

XMRK_XIPMA

ID_XMRK_XIPMA

AC_P13388;

STANDARD;

PRT; 1167 AA.

DT 01-JAN-1990 (Rel. 13, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Melanoma receptor protein-tyrosine kinase precursor (EC 2.7.1.112).
 GN XMRK OR TU.
 OS Xiphophorus maculatus (Southern platyfish).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorphi; Acanthopterygii; Percomorpha; Atherinomorpha;
 OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
 OX NCBI_TaxID=8083;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=90015140; PubMed=2797166;
 RA Wittbrodt J., Adam D., Malitschek B., Maueuer W., Raulf F.,
 RA Telling A., Robertson S.M., Scharf M.;
 RT "Novel putative receptor tyrosine kinase encoded by the melanoma-
 RL inducing Tu locus in Xiphophorus";
 RL Nature 341:415-421(1989).
 RN [2]
 RP REVISION TO 515.
 RA Scharf M.;
 RA Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 RL -!- FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.
 CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.
 CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X16891; CAA34770.2; -;
 DR PIR; S06142; S06142.
 DR HSP; P11362; IFGK.
 DR InterPro; IPR000494; EGFR_L domain.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR002290; Ser_thr_pkinase.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recep_L domain; 2.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00261; FU; 5.
 DR SMART; SM00220; S_TKc; 1.
 DR SMART; SM00219; TyRKc; 1.
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
 DR PROSITE; PS50011; PROTEIN KINASE DOM; 1.
 DR Transmembrane; Glycoprotein; Receptor; Signal; Transferrase;
 KW Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.
 FT SIGNAL 1 25
 FT CHAIN 26 1167
 FT DOMAIN 26 642
 FT TRANSMEM 643 665
 FT DOMAIN 666 1167
 FT DOMAIN 710 977
 FT NP_BIND 716 724
 FT BINDING 743 743
 FT ACT_SITE 835 835
 FT DISULFID 195 204
 FT DISULFID 199 212
 FT DISULFID 228 228
 FT DISULFID 224 236

FT DISULFID 237 245
 FT DISULFID 241 253
 FT DISULFID 256 265
 FT DISULFID 269 296
 FT DISULFID 300 311
 FT DISULFID 315 330
 FT DISULFID 333 337
 FT DISULFID 504 513
 FT DISULFID 508 521
 FT DISULFID 524 533
 FT DISULFID 537 553
 FT DISULFID 556 569
 FT DISULFID 560 577
 FT DISULFID 593 615
 FT DISULFID 618 626
 FT DISULFID 622 634
 FT CARBOHYD 114 114
 FT CARBOHYD 144 144
 FT CARBOHYD 201 201
 FT CARBOHYD 356 356
 FT CARBOHYD 365 365
 FT CARBOHYD 398 398
 FT CARBOHYD 417 417
 FT CARBOHYD 501 501
 FT CARBOHYD 576 576
 FT CARBOHYD 621 621
 SQ SEQUENCE 1167 AA; 129934 MW; 4793E0749DC1D55A CRC64;
 Query Match 37.9%; Score 2569.5; DB 1; Length 1167;
 Best Local Similarity 44.0%; Pred. No. 2e-130;
 Matches 558; Conservative 166; Mismatches 399; Indels 145; Gaps 30;
 QY 4 AALCRGLLIALPPGAASST ---QVCTGDMKRLPASPETHLDMRLHLRYOGQVVOGN 59
 DB 8 AALLQ--LLLVLSISRCSTDPDRKVCQGTSNQTM---LDNHYLKKMKWYSGCNVLEN 62
 QY 60 LEITYLPTNALSFLQDIOEVQGYVLIHANQVPLQRLRIRVGTQLFEDNYALAVLDN 119
 DB 63 LEITYTQENQLSFLQSIQEVGGYVLIAMNEVSTIPLVNLRLIRGQNLVEGNTFLVMSN 122
 QY 120 GDPNNHTTPTVGASPGGLRELQRLSITELIKGVLIQRPOLCYQDTILWKDIFHKNQL 179
 DB 123 YQK-NPSSP--DYVQVGLKQLQLSNLTSLSGGVKVSHPNLLCNVETINMWDIVDKTSNP 179
 QY 180 ALTLIDTNRSRACHPCSPMCKSGRWGSESSDCQSLTRTVCAGC-ARCKGPLPTDCCH 238
 DB 180 TNNLIPAFERQCKDHGCVGSCWAPGPGHCQKFTKLLCAEQCHRCRGPDPIDCNE 239
 QY 239 QCAAGCTGPKHSDCLACLFHFNHSGI CELHCPALVTYNTDTFESMPNPEGYTFGASCVTA 298
 DB 240 HCAGGCTGPRATDCLACRDFNDGTCCKTCTPPPKIYDIVSHQVVDVNPNIKYTFGAACVKE 299
 QY 299 CPYNYLSTDVGSCTLVCPPLHNQVTAEDGTORCEKSKPCARVCYGLGMOYIKANSKFIG 358
 DB 300 CPSNYVYTE--GACVRSCSAGMLEVD--ENGRSKCPDGVCPKVCDDGIGSL--SNTIAVN 356
 QY 359 ITBL-EFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQVFTLEETIYGLVISAAPD 417
 DB 357 STNIRFSNCTKINGDIILNRNSFEGDPHYKIGTMDPEHLWNLTVTVEITGYLVIMWNP 416
 QY 418 SLPDLVSFQNLQVIRGRIHLNGAYS-LTLQGLGISWLGLRLSLRSLRELGSGLALIHHTHLCF 476
 DB 417 NMTLSVLFQNLLEIIRGTRTTFSRGFSFVVQVVRHLQWLGLRLSLKEVSAGNVILKNTLQRY 476
 QY 477 VHTVPMWDLFRNPQALLHTANRPEDECVGBGLACHQLCARGHCWGPGTQCVNCSQFLR 536
 DB 477 ANTNWRRLFRSEDSQSI EYDART-----ENQTCNNECEDGCGWPGTMCVSCSLHVD 529
 QY 537 GQSCVEECVQLQGLPREYVNAHCLPCHPECPQNGSVTCFCGPEADQCVACAHYKDPFPC 596
 DB 530 GGRCVASCNLLQGEPREA QVDRGCVQCHQECLEVTQDLSLTCYGPANCKSAHFQDGPQC 589
 QY 597 VARCPGSKVDPDLSYMPDWKPFPPDEBEGACQPCPINCTHSCVDLDDKGCFAEORASPLTSIVS 656

```
Db 590 IPRCPHGLTGDGTL-IWKYADKMGQCPCHQNCCTQCGSGPLSGCRGD-IVSHSLAVG 647
Qy 657 AVGILLVVLGVVFGVILIKERQOKIRKYWRRLLOETELVEPLTPSGAMNQAMRLK 716
Db 648 LVSGLLTIVALLIVLRRRIK-RKRTIRCLQEKELVEPLTPSGQAPFLRLK 706
Qy 717 ETELKVKVLGSGAFQGVYK-----ANSKFIGITELVRLRENTSPKANKEILDAYVAG 770
Db 707 ETEFKORVLGSGAFVTVYKGLWNPDCENIRIPVAKVLRATSPKVNQEVLDAYVAG 766
Qy 771 VGSYVSRLLGICLTSTVQLVQLMPYGLLDHVRNRLGSLDNLNWCQIAKMSYL 830
Db 767 VDHPCVRLGICLTSAVQLVQLMPYGLLDYVRQHQERICGQWLLNWCQIAKMSYL 826
Qy 831 EDVRLVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGGKVPKIKMALESILR 890
Db 827 EERHLVHRDLAARNVLLKNPNHVKITDGLSKLLTADEKEYQADGGKVPKIKMALESILQ 886
Qy 891 RRFTHQSDVMSYGVTVWELMTFGAKPYDGPAREIPOLLEKGERLPPOPICTIDVYIMV 950
Db 887 WTYTHQSDVMSYGVTVWELMTFGSKPYDGPIDAKIASVLENGERLPPOPICTIEVYIIL 946
Qy 951 KCMWIDSECRPRFRELVSFSEFMARDPQFVVIQNEIDLGPASPLDSTFYRSLLEDMDGD 1010
Db 947 KCMWIDPSSRPRFRELVSFSEFMARDPSRYLVIQ--NLPSLDRRLFSRLSSDD--D 1001
Qy 1011 LVDAEYLVPOGGFFCPDPAPGAGVMVHRHRSSTSGGGDLTLGLLEPSEEAAPRPLA 1070
Db 1002 VVDAEYLVLPYKRI-----NRQGS-----BFCI 1024
Qy 1071 PSEGAGSDVFDGDLGMAAGKLSLTHDPSPLQRYSEDPTV-PLPSETDGYVAPLTCSP 1129
Db 1025 PPTGH-----PVRENSITLRLNSDPTQNALEKDLGH----- 1056
Qy 1130 QPEYVNPQDVRPQ-----PSPRE-----GLP-AARPAGATLERAKTLSPCKNGVVK 1176
Db 1057 --EYVNPQGETSRLSDIYNPNVEDLTDGWPVLSLSSQEAETNFSRPEYLTNQNSL-- 1112
Qy 1177 DVFAFGAVENPEYLTQGGAPQHPPPAPSPAFDNLVYDQDPPERGAPPSTFKGTPT 1236
Db 1113 -PLVSSGSMDDPYD---QAG-----YQAAF-----LPQTGALTGNMGFLPA 1149
Qy 1237 AENPEYLG 1244
Db 1150 AENLEYLG 1157
PRT; 1342 AA.
ID_ERB3 HUMAN STANDARD; PRT; 1342 AA.
AC P21850;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (c-erbB3) (Tyrosine kinase-type cell surface receptor HER3).
GN ERBB3 OR HER3
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90083234; PubMed=2687875;
RA Kraus M.H., Issing W., Miki T., Popescu N.C., Aaronson S.A.;
RT "Isolation and characterization of ERBB3, a third member of the
RT ERBB/epidermal growth factor receptor family: evidence for
RT overexpression in a subset of human mammary tumors.";
RL proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90311312; PubMed=2164210;
```

```
RA Plowman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
RA Todaro G.J., Shoyab M.;
RT "Molecular cloning and expression of an additional epidermal growth
RT factor receptor-related gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
RN [3]
RP SEQUENCE FROM N.A. (SHORT FORM).
RC TISSUE=Placenta;
RX MEDLINE=93282822; PubMed=7685162;
RA Katoh M., Yazaki Y., Sugimura T., Terada M.;
RT "c-erbB3 gene encodes secreted as well as transmembrane receptor
RT tyrosine kinase.";
RL Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC SECRETED (SHORT FORM).
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
CC SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC -1- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M29366; AAA35790.1; -
DR EMBL; M34309; AAA35979.1; -
DR EMBL; S61953; AAB26935.1; -
DR PIR; A36223; A36223.
DR HSP; P11362; LFQK.
DR Genew; HGNC:3431; ERBB3.
DR MIM; 190151; -
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; FALSE_NEG.
DR PROSITE; PSS0011; PROTEIN_KINASE_DOM; 1.
DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Alternative splicing.
FT SIGNAL 1 19
FT CHAIN 20 1342
FT DOMAIN 20 643
FT TRANSMEM 644 664
FT DOMAIN 665 1342
FT DOMAIN 709 966
FT NP_BIND 715 723
FT BIND 742 742
FT ACT_SITE 834 834
FT DISULFID 186 194
FT DISULFID 190 202
POTENTIAL.
RECEPTOR PROTEIN-TYROSINE KINASE ERBB-3.
EXTRACELLULAR (POTENTIAL).
CYTOPLASMIC (POTENTIAL).
PROTEIN KINASE.
ATP (BY SIMILARITY).
ATP (BY SIMILARITY).
BY SIMILARITY.
BY SIMILARITY.
```

```
FT DISULFID 210 218 BY SIMILARITY.
FT DISULFID 214 226 BY SIMILARITY.
FT DISULFID 217 235 BY SIMILARITY.
FT DISULFID 231 243 BY SIMILARITY.
FT DISULFID 246 255 BY SIMILARITY.
FT DISULFID 259 286 BY SIMILARITY.
FT DISULFID 290 301 BY SIMILARITY.
FT DISULFID 305 320 BY SIMILARITY.
FT DISULFID 323 327 BY SIMILARITY.
FT DISULFID 500 509 BY SIMILARITY.
FT DISULFID 504 517 BY SIMILARITY.
FT DISULFID 520 529 BY SIMILARITY.
FT DISULFID 533 549 BY SIMILARITY.
FT DISULFID 552 565 BY SIMILARITY.
FT DISULFID 556 573 BY SIMILARITY.
FT DISULFID 576 585 BY SIMILARITY.
FT DISULFID 589 610 BY SIMILARITY.
FT DISULFID 613 621 BY SIMILARITY.
FT DISULFID 617 629 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 353 353 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 408 408 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 437 437 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 469 469 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 522 522 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 566 566 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 616 616 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPLIC 141 183 EILSGVYIEKNKDLCHMDTIDWRDIAELVVKNGR
SC -> GQFMVPSGLTPQAPQADWYLLDDPRLLTLSASK
VPVTLAAV (IN SHORT ISOFORM).
MISSING (IN SHORT ISOFORM).
E -> G (IN REF. 2).
E -> G (IN REF. 2).
1342 AA; 148097 MW; 7201E7F66CA374BD CRC64;

Query Match 34.2%; Score 2314.5; DB 1; Length 1342;
Best Local Similarity 39.6%; Pred. No. 1.1e-116;
Matches 520; Conservative 191; Mismatches 462; Indels 141; Gaps 35;

QY 10 GLLIALPPGNA--STOVCTGTDMLKRLPASPEHDLMLRLHYOGCVQVGNLELTPLT 67
DB 11 GLLFSLARGSEVGNQAVCPGLTGLSVTGDENQYQTLKLYRCEVVMGNLEIVLTGH 70
QY 68 NASLSFLQDIQEVGYVLIHNOVQVPLORLRIVRGTFQLPEDNYALAVLDNGDPLNNTT 127
DB 71 NADLSFLOWIREVTGYVIVLVANNEFTPLPLNLRVVRGTQVYDGKFAIFVM-----LNYNT 125
QY 128 PVTGASPGGLRELQRLSRLTEILKGGVLIQRNPOLCYQDTILWKDIFPHKNQALTLTIDTN 187
DB 126 ----NSSHALRQLRLTQLTEILSGGVYIEKNKDLCHMDTIDWRDIAELVVKD 178
QY 188 RSRACHPCSPMKCSRWGSESDQSLTRTVACGC-ARCKGBLPTDCHEQCAAGCTG 246
DB 179 NGRSCPPCHEYCKG-RCWGPGESECDTLTKTICAPQCNGHCFGNPNQCCHECAGCGS 237
QY 247 PKHSDCLACLFHNSGICELHCPALVTYNTDTRESMPNPEGRYTFGASCVTACFYNYLST 306
DB 238 PQDTCFACHFNDSGACVPCPCOPLVKNLTFQLEPNHTKYQYGVGVASCCHNFV-V 296
QY 307 DVGSCITLVCPLHNOEVTAEADGTQRCCKSPCARVCYGLGMQYIKANSKF--IGITELE- 363
DB 297 DQTSVCVRACPDKNQEV-D-KNGLKMCPCGGLCPKACEGTG-----SGSRQTVDSSNIDG 350
QY 364 FAGCKKIFGSLAFIPESPDGDPASNTAPLOEQLOVPFETLEEITGYIYISAWPDSLPDLS 423
DB 351 FVNCTKILGNLDFLTILGNDGDPWHKIPALDPEKLNLFVREITGYLNIQSWPPEHMFNS 410
QY 424 VFQNLQVIRGRLHNGAYS-LTLQGLGISWGLRSLRELCSGLALIHNNTHLCPVHTVPW 482
DB 411 VFSNLTTIGRSLYNGRFSLLIMKNLAVNTSLGFSLSKEISAGRIYISANRQLCYHHSLNW 470
```

```
QY 483 DQLFRNPQHALLHTA-NRPEDECVGEGGLACHQLCARGHCWGPPTQCVNCSQFLRGECV 541
DB 471 TKVLRGPTTEERLDIKHNRRDRCDVABGKVCVDCPLCSSGGCWGPQGQCLSRNYSRGVCV 530
QY 542 EECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPDFCVARCP 601
DB 531 THCNFLNGEPREFAEAECSCHPECPMEGTATCNGSGSDTCAQCAHFRDGHPCVSSCP 590
QY 602 SGVKPDLSYMPIWKFDFDEGACQPCINCTHSCVDLDDKGCPEAQR-----SPLTSTVSA 657
DB 591 HGVLG--AKGPIYKPDVQNECRCHENCCTGCKGPELQCLGTLVLKTKHTLTMALTV 648
QY 658 VVGILLVVVLGVVFGILLIKRQOKIR-KYTMRLLOETELVELBPLTPSGAMPNQAQMRILK 716
DB 649 IAG--LVVIFPMMLGGTFLYWRGRRIONKRAMRRLYERGESIEPLDPS-EKANKVLARIFK 705
QY 717 ETELKVKVLGSGAFGOYIK-----ANSKFIGITELVLRENTSPRANKIILDEAYMAG 770
DB 706 ETELKLVKVLGSGVFGTVHKGWIPGESIKIPVICIKVIEDKSGRQSFQAVTDHMLAIGS 765
QY 771 VGSPPYVRLIGICLTSTVQLVTQIMPYGCILLDHVRENRRGLSGQDLINWCMIKAGMSYL 830
DB 766 LDHAHIVRLGLCPGSSQLVTOYPLGSLLDHVROHRRGALGPQLLLNWGVQIAKGMYYL 825
QY 831 EDVRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKMALESILR 890
DB 826 EEHGMVHRNLAAARNVLLKSPQVQVADFGVADLLPDDKQLLYSEAKTPIKMALESIFH 885
QY 891 RRTHOSDVMVSYGVYVWELMTFGAKVDGIPAREIPDLLEKGRLOPPCTIDVYMW 950
DB 886 GKYTHOSDVMVSYGVYVWELMTFGAEPYAGRLAEVDPDLLEKGRLAQPOICTIDVYMWV 945
QY 951 KCMWIDSECRPRRELVSFMRMARDPQRFVVIQNEEDLGA---SPLDSTFYRSLLEDDDD 1007
DB 946 KCMWIDENIRPTKELANEFTRMARDPPRVLVIKRES-GPIAGPEPHGLTNKCLEVE 1004
QY 1008 MGDVDAEYLVPOQGFCCDPAPGAGVMVHRRSRSSSTRSGGDLTLGLEP-SEEAAPR 1066
DB 1005 LEPELDLDDLEAED-----NLATTLTGSALSLPVTGLNRRPGSQ 1045
QY 1067 SPLAPSEGAGSDVFDGLGMAAKQLSLPTH-D-PSPLQRYSEDPTVPLP-----SETD 1119
DB 1046 SLLSPSSGY-MPMNQNLGESCQESAVSGSSERCPRPVSLH-----PMRPGCLASESE 1098
QY 1120 GYVA-----PLTCSPOPE-----YVNPQDVPRPPOPPSPREGP----- 1150
DB 1099 GHTVGEAELEQKVMCRSRSRSPRPRGDSAYHSQRHSLLTPTVPLSPGLEEEDVNG 1158
QY 1151 --LPAARPAGATLERAKTILSP-GKNGVY-----KDVAFGGAVENPEYLTPOGGAAPQ 1201
DB 1159 YVMPDTHLKGTPSSREGTLSSVGLSSVLGTEEDED-----EYEVNRRRRHSP-P 1209
QY 1202 HPPAPSPAFDNLVYWD-----QDPPERGAPSTFKGTPTAENPEYL 1243
DB 1210 HPPRPSLEELGYEYMDVGSLSASLGSTQSCPLHPVIMPTAGTTDDEYEM 1263

RESULT 10
ERB3_RAT
ID ERB3_RAT STANDARD; PRT; 1339 AA.
AC Q62799; Q62955;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (c-erbB3).
GN ERBB3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN 1)
RP SEQUENCE FROM N.A.
```

RC STRAIN=Sprague-Dawley; TISSUE=Liver;
 RX MEDLINE=96096535; PubMed=9522190;
 RA Hellyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Koland J.G.;
 RT "Cloning of the rat ErbB3 cDNA and characterization of the
 RL recombinant protein.";
 RL Gene 165:279-284 (1995).
 [2]
 RN REVISIONS TO 85; 513 AND 565.
 RP Hellyer N.J., Koland J.G.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 [3]
 RN SEQUENCE OF 922-1097 FROM N.A.
 RP STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;
 RC MEDLINE=97184212; PubMed=9030624;
 RA Carroll S.L., Miller M.L., Frohnert P.W., Kim S.S., Corbett J.A.;
 RT "Expression of neuroregulating and their putative receptors, ErbB2 and
 RL ErbB3, is induced during Wallerian degeneration.";
 RL J. Neurosci. 17:1642-1659(1997).
 CC -|- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTK.
 CC -|- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -|- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
 CC (POTENTIAL).
 CC -|- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -|- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
 CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
 CC -|- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
 CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
 CC PHOSPHATIDYLINOSITOL 3-KINASE.
 CC -|- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U29139; AAC28498.2; -;
 DR EMBL; U52530; AAC53050.1; -;
 DR HSSP; P11362; IFGK.
 DR InterPro; IPR000494; EGFR_L domain.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recep_L domain; 2.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00261; FU; 5.
 DR SMART; SM00219; Tyrc; 1.
 DR PROSITE; PS00107; PROTEIN KINASE ATP; FALSE NEG.
 DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
 DR PROSITE; PS00011; PROTEIN KINASE DOM; 1.
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KW Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
 FT SIGNAL 1 19
 FT CHAIN 20 1339
 FT DOMAIN 20 643
 FT TRANSMEM 644 662
 FT DOMAIN 663 1339
 FT DOMAIN 183 259
 FT DOMAIN 707 964
 FT NP_BIND 713 721
 FT BINDING 740 740
 FT ACT_SITE 832 832
 FT DISULFID 186 194
 FT DISULFID 190 202
 FT DISULFID 210 218
 FT DISULFID 224 226
 FT DISULFID 227 235

FT	DISULFID	231	243	BY SIMILARITY.
FT	DISULFID	246	255	BY SIMILARITY.
FT	DISULFID	259	286	BY SIMILARITY.
FT	DISULFID	290	301	BY SIMILARITY.
FT	DISULFID	305	320	BY SIMILARITY.
FT	DISULFID	323	327	BY SIMILARITY.
FT	DISULFID	500	509	BY SIMILARITY.
FT	DISULFID	504	517	BY SIMILARITY.
FT	DISULFID	520	529	BY SIMILARITY.
FT	DISULFID	533	549	BY SIMILARITY.
FT	DISULFID	556	573	BY SIMILARITY.
FT	DISULFID	576	585	BY SIMILARITY.
FT	DISULFID	589	610	BY SIMILARITY.
FT	DISULFID	613	621	BY SIMILARITY.
FT	DISULFID	617	629	BY SIMILARITY.
FT	CARBOHYD	126	126	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	250	250	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	353	353	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	408	408	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	414	414	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	437	437	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	469	469	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	522	522	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	566	566	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	616	616	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CONFLICT	1028	1028	L -> P (IN REF. 3).
SQ	SEQUENCE	1339	AA; 147545	MM; 0AA5F2402BBDFLE CRC64;

Query Match 33.1%; Score 2241.5; DB 1; Length 1339;
 Best Local Similarity 39.8%; Pred. No. 8.6e-113;
 Matches 512; Conservative 168; Mismatches 440; Indels 165; Gaps 36;

QY	3	LAALRWGLLLALLPPGAA---	STQVCTCTDMKRLPASPETHDMLRLHLYQCQVVQGN	59
DB	7	LQVLC----	FLLSLARGSEMGNSQAVCPGPTGLNSVTDGADNQYQTLTKLYKEVVMGN	62
QY	60	LEITLPTNASLFLQDIOEVQGYVLIAHNQVQVPLQRLRIVRGTLQFEDNTALAVLDN	119	
DB	63	LEIVLTGHNADLSFLQWIREVTGYVLVAMNEFSLPLPNLRVVRGTVQVYDGKFAIFW--	120	
QY	120	GDPLNNTPTVGTASPGGLRELQRLSITELKGVILTORNPOLCYQDTILWKDIFHKNNQL	179	
DB	121	---LNTNT---NSSHALROLKFTQLTTELSSGVYIEKNDKLMCHMDTIDWRDVRVR--	170	
QY	180	ALTLDITNRSRACHPCSPMKSGRCSGSESDCSLRTVTCAGGC-ARCKGPLPTDCCHE	238	
DB	171	GAEIVVKNNGANCPPCHEVCCKG-RCWGPDPDDCQILTITKTCAPQCNGRCFNPNOCCHD	229	
QY	239	QCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTA	298	
DB	230	ECAGGCGSPQDTCDFACRRFNDSDGACVPRCPPLVYNKLTFLQLEPNHTKYQYGGVCVAS	289	
QY	299	CPYNYLSTDVGSCTLVPLHNEVTABDTQRCCKSKCPARVCYGL--GMOYIKANSKF	356	
DB	290	CPHNFV-VDQTFVCRACPPDKMEVD-KHGLKMCPCGGLCPKACEGSGSRQYTVDDSN	347	
QY	357	IGITELEFACCKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFEETLEETGVLYISAWP	416	
DB	348	ID----GFVNCTKILGNLFLITGLNVDPMHKIPALDPEKLVNFRVTRVITGLYNTQSWP	403	
QY	417	DSLPLDSVFONLQVIRGRILHNGAYS-LTIQIGISWLGRLSLRLSGSLALHNNTHLC	475	
DB	404	PHMNFVSFSLNTTIGRSLYNGFSLIMKNLVTSLSGFRSLKELISAGRVYISANQQLC	463	
QY	476	FVHTVPMDQLFRNPHOALLHTA-NRPEBCVGEGLACHOLCARGHCWGPGPTQCVNCSQF	534	
DB	464	YHSLNWTLLRGPSEERLDIKYDRPLGECLABGKVCPLCSSGGCGWPGPGQCLSCRNY	523	
QY	535	LRQCECCEBRCVLOGLPREVYNARHCLPCHPECPQNGSVTCFPGPADQCVACHYKDP	594	
DB	524	SREGVCVTHCNFLQGBEFREFVHEAQCFSCHPPELPMEGTSTCNGSGSDACARCAHFRDGP	583	
QY	595	FCVARCPGKVPDLSYMPIWKFPDDEGACQPCPCINCTHSC--VDLDDKGPAPQASPLT	652	

Db 584 HCVNSCPHGLG--AKGIYKYPDAQNECRPCHECHTQGCNGPELQDCLGQAEVLMSPKH 641
Qy 653 SIVSAVVGILVAVGVVFGILIKRRQOKIR-KYTMERLLQETELVELPTPSGAMPNOAQ 711
Db 642 LVIAVTVG--LAVILMILGGSFLYWRGRIONKRAMRYLGRGESIFLPDS- EKANKVL 698
Qy 712 MRILKETELRKVKVLGSGAFQYTK-----ANSKFITGITELVURENTSPKANKEILDEA 765
Db 699 ARIFKETELRKVLGSGVFGVTHKGIWIPGESIKIPVCIKVIEDKSGRQSQFQAVTDMH 758
Qy 766 YVMAGVSPVYSRLIGLCTSTVOLTPMYPGCLLDHVENRGLSGODLLANCMQOIAK 825
Db 759 LAVGSLDHAHIVRLGLCPGSSGLVQYLPGLGSLLDHVKOHRRTGLPQLLNWGVQIAK 818
Qy 826 GMSYDELVRHLDAARNVLKSPNVKIKTDFGLARLLDIDETEHYHADGKGVPIKMMAL 885
Db 819 GMYLEHSMVHRDLARNVLMKSPSQVQVADFGVADLLPDDKQLLHSEAKTPIKMMAL 878
Qy 886 ESILRRFTHQSDVMSYGVTVWELMTGAKPYDGIPIAREIPDLLEKGERLPQPPICTIDV 945
Db 879 ESIFGKYTHQSDVMSYGVTVWELMTGABPYAGLRLAEIPDLLEKGERLAQPOICTIDV 938
Qy 946 YMIWVKWMDISECRPRFELVSFSESRWARDPQRFVQIONEDLGPASPLDSTFFVRSILED 1005
Db 939 YMWVKWMDIDENIRPTFKELANEFTRWARDPPRYLVIKRAS-GPGTP--PAAEPSVLTT 995
Qy 1006 DMDGLVDABEYLVPQGFPCDPAPGAGGVVHRRSSSTRSGGDLTLGLEPSES--- 1062
Db 996 KEL-----QEAELEPEL-----DLDLLEAEESGLA 1021
Qy 1063 -----EAPRSIAPSEG-----AGSDVFDGLGMGAAGLQSLPT 1097
Db 1022 TSLGSALSLPTGLTRPRGSSLLSPSSGYMPNMQSSILGEACLDASVILGREGQPSRISL 1081
Qy 1098 HDPSPLQYSEDPTVPLPSETDGVV----APL-----TC-----SPOPE-----YVNPQ 1137
Db 1082 H-PIPRGR-----PASESEGHVTSGEAELOEKVSVCRSRSRSPRPGDSAYHSQR 1133
Qy 1138 DVRPQPSPREGP-----LPAARPAGATLERAKTLP-SGKGVV-----KDVF 1179
Db 1134 HSLLTPTVPLSPGLEBEDGNGYVMPDTHLRGASSREGTLSSVGLSVLGTBEED--- 1191
Qy 1180 AFGGAVENPEYLTPOGGAAPQHPPP 1204
Db 1192 -----EEYEVYMKRKGSP-PRPP 1209

RESULT 11

ID EGFR DROME STANDARD; PRT; 1426 AA.
AC P0412; Q61601; Q9W2G0; P81868;
DT 13-AUG-1987 (Rel. 05, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Egfr)
DE (Gurken receptor) (Torpedo protein) (Drosophila relative of ERBB).
GN EGFR OR TOP OR C-ERBB OR DER OR CG10079.
OS Drosophila melanogaster (Fruit fly)
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
[1]
RP SEQUENCE FROM N.A. (ISOFORMS TYPES I AND II).
RX MEDLINE=94350209; PubMed=6070664;
RA Clifford R., Schubach T.;
RT "Molecular analysis of the Drosophila EGF receptor homolog reveals
RT that several genetically defined classes of alleles cluster in
RT subdomains of the receptor protein.";
RL Genetics 137:531-550(1994).
RN [2]
RP REVISTONS.

RA Clifford R., Schubach T.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=85124611; PubMed=2982499;
RA Livneh E., Glazer L., Segal D., Schlessinger J., Shilo B.-Z.;
RT "The Drosophila EGF receptor gene homolog: conservation of both
RT hormone binding and kinase domains.";
RL Cell 40:599-607(1985).
RN [4]
RP SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RC STRAIN=Oregon-R, TISSUE=Embryo;
RX MEDLINE=87002474; PubMed=3093080;
RA Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;
RT "Alternative 5' exons and tissue-specific expression of the
RT Drosophila EGF receptor homolog transcripts.";
RL Cell 46:11091-1101(1986).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION ANALYSIS.
RX MEDLINE=99102120; PubMed=9882502;
RA Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;
RT "Several levels of EGF receptor signaling during photoreceptor
RT specification in wild-type, Ellipse, and null mutant Drosophila.";
RL Dev. Biol. 205:129-144(1999).
RN [6]
RP SEQUENCE FROM N.A. (ISOFORM TYPE I).
RC STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Anantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos F.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fogle C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Sidenkian I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [7]
RP SEQUENCE OF 959-1078 FROM N.A.
RC STRAIN=Daekwanryeong;
RX MEDLINE=85137938; PubMed=2983232;
RA Wadsworth S.C., Vincent W.S. III, Bilodeau-Wentworth D.;
RT "A Drosophila genomic sequence with homology to human epidermal


```

Db 603 IADCGYISNAYK--FDNRKTKICHPCKR-----TCNGAGADHCOEQRVHVRDGHQVSEC 654
Qy 601 P-----SGVK-----PDL 608
Db 655 PKNYNDRGVCRECHATCGTGKPTIGIGACTTCNLAINNDAIVKRCCLKDKCPD- 713
Qy 609 SYMFIWKF--PDEGACQP-----CPI-----NCTH-----632
Db 714 GY--FWYVHPQEOGSLKPLAGRAVCRKCHPLCLCTNYGYHEQVCSKCTHYKRREQCET 771
Qy 633 -----SC-----VLDLDDG-----641
Db 772 ECPADHTDEBQRCFORHPCCNCTGPGADDDCKSRNFKLFANETGPGYVNMFMNCTS 831
Qy 642 -CPAEOR-----ASPLTS-----IVSAVVGILLVVLGVVFGI 673
Db 832 KCPLMRHVNQYTAIGPYCAASPRSSKITANLDVNMFIITGAVLVPTICILCVV--T 889
Qy 674 LKRRQOKIRKYT--MRRLQOETELVEPLTPSGAMPNQAOQMRILKETELRKVKVLGSGAF 731
Db 890 YICRQOKAKETVKMTMALSGCDSBPLRPSNIGANLCKLRIVKDAELRKGGLGNGAF 949
Qy 732 GOYIKA-----NSKF-IGITELVLRENTSPKANKELDEAYVMAGVGSPPYVSRLLGIC 783
Db 950 GRVYGVWVPEGENVKIPVAIKELL--KSTGAESSEBFLREAYIMASEEHNLLKLAVC 1007
Qy 784 LTSTVQLVTLQMPYGCCLLDHVRNRRGLSQDLNWCWQAKGMSYLEDLVRLVHRDLAAR 843
Db 1008 MSSQMLLITQLMPLGCLLDYVNRNRDKIGSKALLNWSQTAKGMSYLEEKLVRDLAAR 1067
Qy 844 NVLKSNNHVKITDFGLARLLDIDETEHADGGKVPKMKMALESILRRRTHQSDVMSYG 903
Db 1068 NVLVQTPSLVKITDFGLAKLLSSDSNEYKAAGGKWPVKWLALEICIRNRVFTSKSDVWAF 1127
Qy 904 VTMELMTFCAKPYDGIPIAREIPDLLEKGERLPPOPICTIDVYMWKMWIDSECPRF 963
Db 1128 VTIWELLTUGQRPHENIPAKDIPLIEVGLKLEQPEICSDIYCTLLSCWHLDAAMRPTF 1187
Qy 964 RELVSEFSRMARDPQRFVWQNEPLG--PASPLDSTFVRSLLDD--DMGDLVDAREYL 1018
Db 1188 KQLTTFVAFEPARDGRYLAIPGDKFTRLPA-----YTSQDEKDLRKLAPTIDGSEAI 1240
Qy 1019 VPQGGFCPPDAPAGAGGWHHRSSSTRGGGDLTLGLEPSEFAP-----RSPLAPS 1072
Db 1241 AKPDVYLQPKAAPGPS-----HRTDCT-----DEMPKLNRYCKDPSKN 1279
Qy 1073 EGAGSDVFDG---DLGMAAGKLQSLTHDPSPLQRYSEDPVLPSETDGYVAPLTCSP 1129
Db 1280 SSTGDDERDSSAREVGVGNLR-----LDLPVDEDDYLMPTCPQ 1317
Qy 1130 QPEYVNPQDVRPQPPSGPLPAAPAGATLERAKTLPCKNGVVKDVFAGFAGVNP 1189
Db 1318 GPNNNNMN-----NPNQNNMAVGVAGYM-----DLTGVPVSDNPE 1356
Qy 1190 YL-----TPQGAAPQPH-----PPPAFSP-AFDNLYYWD 1218
Db 1357 YLLNAQTLGVGESPIQTIGIPWGGPGTMEVKVPMFPGSEPTSSDHEYND 1408

RESULT 12
ID ERBB ALV STANDARD; PRT; 634 AA.
AC P00534;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian leukosis virus.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11964;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE=85228222; PubMed=2988784;
RA Nilssen T.W., Maroney P.A., Goodman R.G., Rottman F.M.,
RA Crittenden L.B., Raines M.A., Kung H.-J.,
RT "c-erbB activation in ALV-induced erythroblastosis: novel RNA
RT processing and promoter insertion result in expression of an
RL amino-truncated EGF receptor.";
CC Cell 41:719-726 (1985).
CC -|- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -|- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC IN CHICKS WHICH HAVE BEEN INJECTED WITH THE AVIAN LEUKOSIS VIRUS
CC AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC -|- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC PROTEIN.
CC -|- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M10066; AAA48763.1; ALT_INIT.
DR PIR; A00643; TVCHLV.
DR PIR; B00643; TVFVLV.
DR HSP; P11362; IFGK.
DR InterPro; IPR000719; Euk_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
DR PROSITE; PS00111; PROTEIN KINASE DOM; 1.
KW Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399 PROTEIN KINASE.
FT NP BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT SITE 257 257 BY SIMILARITY.
SQ SEQUENCE 634 AA; 70891 MW; E705E33A0BE01FCC CRC64;

Query Match 24.1%; Score 1633.5; DB 1; Length 634;
Best Local Similarity 50.2%; Pred. No. 1.4e-80;
Matches 356; Conservative 79; Mismatches 141; Indels 133; Gaps 20;

Qy 587 CAHYKDPFCVACPSGVKPDLSYMPFWKFPDEGACQPCINCHSCVDLDDKGPAAEQ 646
Db 3 CAHFDGPHCVKACPAVLGENDTL-VWKYADANAVCOLCHPNCTRGCKGPGLEGCP--- 58
Qy 647 RASPLTSIVSAVV-GILLVVLGVWFGILIKRQOKIRKVTMRLLQOETELVEPLTPSGA 705
Db 59 NSKTSIAAGVVGGLLCLVVGVLGIGLYLRRR-HIVRKTLRRLLOEQLVEPLTPSGE 117
Qy 706 MPNQAOQMRILKETELRKVKVLGSGAFQ-----YKANSKF-----IGITELVLRENTSPKA 757
Db 118 APNQAHRLILKETEFKKVKVLGSGAFGTVVKGLWIPEGEKVKIPVAIKE--LREATSPKA 175
Qy 758 NKEILDEAYVMAGVGSPPYVSRLLGICLTSTVQLVTLQMPYGCCLLDHVRNRRGLSQDLL 817
Db 176 NKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQMPYGCCLLDVIREHKDNIGSQYLL 235
Qy 818 NWCWQIAKMSYLEDLVRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGK 877
Db 236 NWCVQIAKGMNLTLEERLVHRDLAARNLVKTPQHVKITDFGLAKLGLGADEKEYHAEGK 295
Qy 878 VPIKWMALLESILRRRTHQSDVMSYGVTVWELMTFCAKPYDGIPIAREIPDLLEKGERLPQ 937

```



```
Db 296 VPIKMALESLHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIAPASBISSVLEKGERLPQ 355
Qy 938 PPICTIDVYIMVKCWMIDSCRRFRRELVSERFARMARDPQRFVVIQ-NEDLGASPLDS 996
Db 356 PPICTIDVYIMVKCWMIDSCRRFRRELVSERFARMARDPQRFVVIQ-NEDLGASPLDS 415
Qy 997 TFYRSLLDDMDGLVDAEYLVPOQGFCCDPAPGAGGMVHRRHSSTRSGGDLTLG 1056
Db 416 KFYTLMEEDMEDIVDAEYLVPHQGF-----NPSST----- 449
Qy 1057 LEPSEEEAPRSPL-----APSEGAGSDVFDGDLGMAAKGLQSLFTHDPSPLOQYSBDPT 1111
Db 450 -----SRTPLLSLSATSNNATNCID-----RNGQGHVPREDSPVQYSSDPT 493
Qy 1112 VPLPSET--DGYVAPLTCPOPEVYNQPDVPQPPSPREGPLPAARFAGATLERAKTLSP 1169
Db 494 GNFLSEIDDGFL-----PAPEYVQ--LMPKXPS-----TAM 524
Qy 1170 GKNGVWQDF-----AFGGAIVENPEYLTPOGGAAPQPPPPAPSPAFDNL 1214
Db 525 VQNGIYNNISITAIKULPMSRYQNSHSTAVDNEYL-----NTNQSPLAKTVPFESS 576
Qy 1215 YYWQDQ-----DPPE-----RGAPPSTFKGTPTAENPEYLGLDVP 1248
Db 577 PYMTQSGNHQINLNDPQYQDFLNETKPNGLLKVPAENPEYLRVAAP 625

RESULT 13
EGFR CHICK
ID EGFR CHICK STANDARD; PRT; 703 AA.
AC P13387;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER)
DE (Fragment).
GN EGFR.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88261272; PubMed=3260329;
RA Lax I., Johnson A., Hawk R., Sap J., Bellot F., Winkler M.,
RA Ullrich A., Vennstrom B., Schlessinger J., Givol D.;
RT "Chicken epidermal growth factor (EGF) receptor: cDNA cloning,
RT expression in mouse cells, and differential binding of EGF and
RT transforming growth factor alpha";
RL Mol. Cell. Biol. 8:1970-1978(1988).
CC -!- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC dimerization, internalization of the EGF-receptor complex,
CC induction of the tyrosine kinase activity, stimulation of cell DNA
CC synthesis, and cell proliferation.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M20286; AAA48760.1; -.
```

```
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR SMART; SM00261; FU; 4.
DR PROSITE; PS00107; PROTEIN KINASE ATP; PARTIAL.
DR PROSITE; PS00109; PROTEIN KINASE TYR; PARTIAL.
DR PROSITE; PS00011; PROTEIN KINASE_DOM; PARTIAL.
DR Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1..30
FT CHAIN 31..703 EXTRACELLULAR GROWTH FACTOR RECEPTOR.
FT DOMAIN 31..654 POTENTIAL.
FT TRANSMEM 655..667 POTENTIAL..
FT DOMAIN 668..703 CYTOPLASMIC (POTENTIAL).
FT DISULFID 201..214 BY SIMILARITY.
FT DISULFID 222..230 BY SIMILARITY.
FT DISULFID 226..238 BY SIMILARITY.
FT DISULFID 239..247 BY SIMILARITY.
FT DISULFID 243..255 BY SIMILARITY.
FT DISULFID 258..267 BY SIMILARITY.
FT DISULFID 271..298 BY SIMILARITY.
FT DISULFID 302..314 BY SIMILARITY.
FT DISULFID 318..333 BY SIMILARITY.
FT DISULFID 336..340 BY SIMILARITY.
FT DISULFID 513..522 BY SIMILARITY.
FT DISULFID 517..530 BY SIMILARITY.
FT DISULFID 533..542 BY SIMILARITY.
FT DISULFID 546..562 BY SIMILARITY.
FT DISULFID 565..581 BY SIMILARITY.
FT DISULFID 569..589 BY SIMILARITY.
FT DISULFID 592..601 BY SIMILARITY.
FT DISULFID 605..627 BY SIMILARITY.
FT DISULFID 630..638 BY SIMILARITY.
FT DISULFID 634..646 BY SIMILARITY.
FT CARBOHYD 134..134 N-LINKED (GLCNAC.. ) (POTENTIAL).
FT CARBOHYD 190..190 N-LINKED (GLCNAC.. ) (POTENTIAL).
FT CARBOHYD 200..200 N-LINKED (GLCNAC.. ) (POTENTIAL).
FT CARBOHYD 359..359 N-LINKED (GLCNAC.. ) (POTENTIAL).
FT CARBOHYD 368..368 N-LINKED (GLCNAC.. ) (POTENTIAL).
FT CARBOHYD 420..420 N-LINKED (GLCNAC.. ) (POTENTIAL).
FT CARBOHYD 573..573 N-LINKED (GLCNAC.. ) (POTENTIAL).
FT CARBOHYD 578..578 N-LINKED (GLCNAC.. ) (POTENTIAL).
FT CARBOHYD 613..613 N-LINKED (GLCNAC.. ) (POTENTIAL).
FT CARBOHYD 633..633 N-LINKED (GLCNAC.. ) (POTENTIAL).
FT CARBOHYD 648..648 N-LINKED (GLCNAC.. ) (POTENTIAL).
FT NON_TER 703..703
SQ SEQUENCE 703 AA; 77427 MW; AFF2DE11B735A690 CRC64;

Query Match 23.5%; Score 1595; DB 1; Length 703;
Best Local Similarity 44.6%; Pred. No. 1.8e-78;
Matches 316; Conservative 111; Mismatches 251; Indels 30; Gaps 14;

Qy 8 RWGLLALLPPGAA-----STOVCTGTDMLRPLASPETHDLRLHYGCGVQGNLE 61
Db 13 RGAIVLLLLGLVALCSAVEERKVCQGTNNKLTQLGHVEDHFTSLQRMNNCEVLISNLE 72
Qy 62 LTYLPTNASLFLQDIQEYGVYVLIHNOVRQVPLQRIRIVRGTLQFEDNYALAVLDNGD 121
Db 73 ITVEHNRDLTFLKLTQEVAGYVLIHNMWDVIPLENLQIRGNVLYDSFALAVLSNYH 132
Qy 122 PLNNTPTVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILKWDIFHNNQAL 181
Db 133 -MNRKTQ-----GLRELPMKSLSEILNGGVKISNNPKLCNNMDITVLWNDIIDTSRK-PL 182
Qy 182 TLID-TNRSRACHPCSPMKGSRCEWSESSDCOSLRTTVCAGGCA-RCKGPLPTDCCHEQ 239
Db 183 TVLDFAFNSSSPKCPNCTEDHCWAGBONCOTLTKVICAQOQSCRCRCRQVSDCCCHNQ 242
Qy 240 CAAGCTGPKHSDCLACLHFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTAC 299
```

```
Db 243 CAAGCTGPRSDCLACRFRDADTKDTCPLVLYNPTTYQMDVNPBGKYSFGATCVREC 302
Qy 300 PYNLYSTDVSGCTLVCPHLHQEVTAEDGTORCEKCKPCARVCVGLGMQVYKANSKFIGI 359
Db 303 PHNVVTDHSGVRSQNTDITYEV-EENGVRKCKKCDGLCKVCGNGIGIGELKGLS-INA 360
Qy 360 TELE-PAGCKKIFGSLAFPESEFDGDPASNTAPLOPELOLVFETLEITGYLYISAWPDS 418
Db 361 TNIDSPKNCCKINGDVSILPVAFGLDAFTKTLPLDKKLDVFTVKEISGFLLIQAWPDN 420
Qy 419 LPDLVSFONLOVIRGILHNGAVSLTQIGISWLGSLRSLRELGLALHNNTHLCFVH 478
Db 421 ATDLYAPENLEIRGTRKQHQVSLAVNLKIOSLGRSLKEISDGDIAIMKNKLCYAD 480
Qy 479 TVPWDQLFRNPHQALLHTANRPECEVGEGLACHQLCARGHCWGPPTQCVNCSQFRLGQ 538
Db 481 TMNWSLFAFQSKTKIQRNKNNDCTADRHVCDPLSCDVCWCGPFGFHCFSRQK 540
Qy 539 ECVEEVRVQLGPREYVYNAHCLPCHPECPQNG---SVTCFGEADQCACAHYKDPFF 595
Db 541 ECVKQCNILOGEPRERFDSKCLPCHSECLVQNSTAYNTTCSGPDHCKMCAHFIDGPH 600
Qy 596 CVARCPGKPDLSYMPWKFPDEEGACQPCINCHSCVDLDDKGCFAQRASPLTSIV 655
Db 601 CVKACPAGVLGENDTL-VWKYADANAVCQLCHPNCTRGCKPGLEGCP---NGSKTPSIA 656
Qy 656 SAVV-GILLVVLGVWFGIILKRQKIRKYTMRLLOETELVEPLTP 702
Db 657 AGVVGGLLVVLGVGLGLYLRR-HIVRKRTLRLQLQERLVEPLTP 703

RESULT 14
ERBB AVIER
ID ERBB AVIER STANDARD; PRT; 604 AA.
AC P00535;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ES4).
OC Viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=79685;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H;
RX MEDLINE=84026539; PubMed=6313229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erbB gene of avian erythroblastosis virus is a member of the src
gene family."
RL Cell 35:71-78(1983).
RN [2]
RP SEQUENCE OF 1-152 FROM N.A.
RX MEDLINE=84231957; PubMed=6328658;
RA Debure B., Henry C., Benaisa M., Biserte G., Claverie J.-M.,
RT Saule S., Martin P., Stehelin D.;
RT "Sequencing the erbA gene of avian erythroblastosis virus reveals a
new type of oncogene."
RL Science 224:1456-1459(1984).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
tyrosine phosphate.
CC -1- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
IN CHICKENS.
CC -1- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
```

```
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; K02006; AAA42394.1; ALT_INIT.
DR EMBL; K01216; AAA42400.1; -.
DR PIR; A00644; TVYUH.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
DR TRANSFERASE; PS00111; PROTEIN KINASE DOM; 1.
KW Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399 PROTEIN KINASE.
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
FT CONFLICT 29 29 R -> W (IN REF. 2).
FT CONFLICT 140 140 S -> F (IN REF. 2).
FT CONFLICT 146 146 I -> V (IN REF. 2).
FT SEQUENCE 604 AA; 67633 MW; 76EBCDD06745D609 CRC64;

Query Match 23.4%; Score 1588; DB 1; Length 604;
Best Local Similarity 50.1%; Pred.No.3.6e-78;
Matches 347; Conservative 75; Mismatches 134; Indels 136; Gaps 19;

Qy 587 CAHYKDPPEFCVARCPGKPDLSYMPWKFPDEEGACQPCINCHSCVDLDDKGCFAEQ 646
Db 3 CAHFIDGPHCVKACPAGVLGENDTL-VRKYADANAVCQLCHPNCTRGCKPGLEGCP--- 58
Qy 647 RASPLTSIVSNV-GILLVVLGVWFGIILKRQKIRKYTMRLLOETELVEPLTPSGA 705
Db 59 NGSKTPSIAAGVVGGLLVVLGVGLGLYLRR-HIVRKRTLRLQLQERLVEPLTPSGE 117
Qy 706 MPNQAQMRILKETLRKVKVLGSGAFGQ---YIKANSKF---IGITELVLRENTSPPKA 757
Db 118 APNQAHLRIKETEFPKKVKVLGSGAGFTYKGLWIPEGEKVKIPVAIKE--LREATSPKA 175
Qy 758 NKEILDEAYNAGVSGPYVSRLLIGICLTSTVQLVTQIMPYGCLLDHVRENRLGSQLL 817
Db 176 NKEILDEAYMASVDNPHVCRLLIGICLTSTVQLITQIMPYGCLLDYIREHKDNIGSQLL 235
Qy 818 NWCWQIAKMSYLEDLVRLVHRDLAARNVLKSPNHNKVTDFGLARLLDIDETEHADGGK 877
Db 236 NWCVQIAKGMVYLEERRLVHRDLAARNVLKTPQHVKITDFGLAKLGLADEKEYHAEGGK 295
Qy 878 VPIKWMALLESILRRRTHQSDVMSYGVTVWELMTFCAPYDGIIPAREIPDLLEKGERLPQ 937
Db 296 VPIKWMALLESILRIYTHQSDVMSYGVTVWELMTFGSKPYDGIIPASEISVLEKGERLPQ 355
Qy 938 PPICTIDVYMWKCMIDSECRPRELVSFSESRMARDPQRFVWVQI-NEDLCPASPLDS 996
Db 356 PPICTIDVYMWKCMWIDADSPKFEELIAEFESKWARDPPRVLYVQGDREHMLPFTDS 415
Qy 997 TFVRSLLDDDDMGDLVDAEYLVPOQGFPCPDAPAGAGGMVHHRRSSSSRSGGDLTLG 1056
Db 416 KFYRTLMEEDVEDIVDAEYLVPHQGF-----NSPST----- 449
Qy 1057 LEPSSEEAERSPL-----APSEAGSDVFDGLGMGAAGLQSLPHTHDSPLQYSEDP 1111
Db 450 -----SRTPLLSLSLSSATNSATNCID-----RNGQGHVPVRESFVORYSDPT 493
Qy 1112 VPLPSET--DGVVAPLTCSFQPEYNQPDVRPOPSPREGPLPAARPAGATLERAKTSLP 1169
Db 494 GNPLESIDDGL-----PAPEYVQ--LMPKKPSTAM----- 524
Qy 1170 GKNGVVKDVFAP-----GGAVENPEYLTPOGGAAPQPHPPAFSPA 1210
||:|||||
||:|||||
```

```
Db 525 ----VQOQVNFISLTAISKLPMSDRYQNSHSTAVDNPEYL-----NTNQSLAKTV 572
Qy 1211 FDNLVYNDODPPERCAPSTFKGTPTAENPEY 1242
Db 573 FESSPYWTQSCNHQ-----INLNDPDY 594

RESULT 15
ERBB AVIEU
ID -ERBB AVIEU STANDARD; PRT; 540 AA.
AC P11273;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbb (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ts167).
OC Viruses; Retroviral viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=103898;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87064456; PubMed=2878364;
RA Choi O.R., Traitor C., Graf T., Beug H., Engel J.D.;
RT "A single amino acid substitution in v-erbB confers a thermolabile
RT phenotype to ts167 avian erythroblastosis virus-transformed erythroid
RT cells."
RL Mol. Cell. Biol. 6:1751-1759(1986).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M13179; AAA2401.1; -
CC PIR; A25231; TVFEB.
CC HSP; P11362; IFCK.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR001245; Tyr_pkinase.
CC Pfam; PF00069; pkinase; 1
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00219; TyrKc; 1.
CC PROSITE; PS00107; PROTEIN KINASE ATP; 1.
CC PROSITE; PS00109; PROTEIN KINASE TYR; 1.
CC PROSITE; PS50011; PROTEIN KINASE DOM; 1.
CC Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
CC Glycoprotein; Phosphorylation.
FT DOMAIN 132 399 PROTEIN KINASE.
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
FT VARIANT 270 270 H -> D (IN THERMOLABILE V-ERBB).
SQ SEQUENCE 540 AA; 60412 MW; 5B53297AA068B65D CRC64;

Query Match 22.4%; Score 1520; DB 1; Length 540;
Best Local Similarity 52.7%; Pred. No. 1.4e-74;
Matches 326; Conservative 68; Mismatches 129; Indels 96; Gaps 15;

Qy 587 CAHYKDPFVCARCPGVKPDLSYMPIWKFPDEGACQPCPINCTHSCVDLDDKGCBAEQ 646
Db 3 CAHFIDGPHCVKACPAVLGENDTL-VMKYADANAVCQLCHPNCTRGCGLEGCP--- 58
Qy 647 RASPLTSIVSAVV-GILLVVVGVFGILIKRQOKIRKQMRLLQETELVEPLTPSGA 705
Db 59 NGSKTPSIAGVCGGLICLVVVGIGLYLRRR-HIVKRTLRLQLRELVEPLTPSGE 117
Qy 706 MPNOAQWRILKETELRKVKVLGSGAFQGYIK-----ANSKFIGITELVLENTSPKANK 759
```

```
Db 118 APNOAHLRIKETEFKKVKVLGFGAGFTVYKGLWIPEGEKVTIPVAIKELREATSPKANK 177
Qy 760 EILDEAYVMAGVCSPVVRLIGICLTSTVQLTQMLPYGCLLDHVRNCRGLGSQDLLNW 819
Db 178 EILDEAYVMASVDNPHVCRLLGICLTSTVQLITQMLPYGCLLDYIREHKDNIGSQYLLNW 237
Qy 820 CMQIAKMSYLEDVRLVHRDLAARNVLKSPNHNKITDFGLARLLDIDETEHADGGKVP 879
Db 238 CVQIAKGMVLEERHVMVHRDLAARNVLVKTPOHVKITDFGLAKQLGADEKYEHAEGKVP 297
Qy 880 IKWMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPP 939
Db 298 IKWMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGPASEISSVLEKGERLPQPP 357
Qy 940 ICTIDVYIMVCKWMDSECRPRFRELVSFFSRMARDPQRFVVIQ-NEDLGPASPLDSTF 998
Db 358 ICTIDVYIMVCKWMSDADSRPKRELIAEFSSKARDPPRYLVIQGDHMLPSPTDSKF 417
Qy 999 YRSLLDDMDGLVDAEYLVPOQGFPCPDAPGAGMGVHRRHRSSTRSGGDLTLGLE 1058
Db 418 YRTLMEEDMEDIVDADEVILVPHQGF-----NSPST----- 449
Qy 1059 PSEEEAPRSP-----APSEGAGSDVFDGDLGMAAKGLQSLPHTDPSPLQRYSEDPSTVP 1113
Db 450 -----SRTPLLSSLSATSNNNSATNCIDRNGG-----H----- 476
Qy 1114 LPSETDGYVAPLTCSPQPEYVNPQDVPYRPPSPREGPLPAARPAGAT-LERAKTLSPGKN 1172
Db 477 -PVREDGFL-----PAPEYVNO--LMPKKPSTAMVQNIYVISTAKLPIDSRYON 527
Qy 1173 GVVKDVFAGGAVENPEYL 1191
Db 528 -----SHSTAVDNPEYL 539

Search completed: July 22, 2003, 09:18:49
Job time : 16.5911 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 09:01:42 ; Search time 46.6008 Seconds
(without alignments)
5522.503 Million cell updates/sec

Title: SEQ4-710-730-12
Perfect score: 6775
Sequence: 1 MEALALCRWGLLLALLPPGA.....TFKGTPTAENPEYGLDVPV 1249

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 21: *
1: sp_archaea: *
2: sp_bacteria: *
3: sp_fungi: *
4: sp_human: *
5: sp_invertebrate: *
6: sp_mammal: *
7: sp_mhc: *
8: sp_organelle: *
9: sp_phage: *
10: sp_plant: *
11: sp_rodent: *
12: sp_virus: *
13: sp_vertebrate: *
14: sp_unclassified: *
15: sp_ivirius: *
16: sp_bacteriap: *
17: sp_archaeap: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	6091	89.9	1259	6 O18735	O18735 canis famil
2	3028	44.7	1209	11 Q9QX70	Q9QX70 rattus norv
3	2999	44.3	1210	11 Q9EP98	Q9EP98 mus musculu
4	2611	38.5	1165	13 Q9YH40	Q9YH40 xiphophorus
5	2577.5	38.0	1137	13 Q9W6F6	Q9W6F6 gallus gall
6	2191	32.3	1328	13 P79754	P79754 fugu rubrip
7	1930.5	28.5	1433	5 Q9BIH9	Q9BIH9 anopheles g
8	1871	27.6	419	4 Q9UK79	Q9UK79 homo sapien
9	1739	25.7	367	11 Q8R2X1	Q8R2X1 mus musculu
10	1697.5	25.1	412	4 Q8WYV0	Q8WYV0 homo sapien
11	1604	23.7	729	15 Q86712	Q86712 avian rous-
12	1602	23.6	567	15 Q86714	Q86714 avian rous-
13	1543.5	22.8	962	15 Q64895	Q64895 avian eryth
14	1535	22.7	545	15 Q85468	Q85468 avian eryth
15	1506.5	22.2	655	11 Q9WVF5	Q9WVF5 mus musculu
16	1490.5	22.0	643	11 Q9ERV6	Q9ERV6 mus musculu

```

17 1206 17.8 1193 5 Q9YIX8
18 1126 16.6 527 13 Q90836
19 1125.5 16.6 1368 5 Q23821
20 1109 16.4 1717 5 Q26566
21 1001.5 14.8 478 11 Q9ESE0
22 942.5 13.9 599 13 Q9PSH2
23 906 13.4 165 4 Q14256
24 887 13.1 176 11 Q923V5
25 806.5 11.9 346 13 P11776
26 778 11.5 435 5 Q8SZW1
27 754.5 11.1 311 13 Q99162
28 734 10.8 331 4 Q9BUD7
29 723 10.7 149 6 Q9B666
30 698.5 10.3 1362 13 Q9PVZ4
31 682 10.1 1671 5 Q9NJV5
32 667.5 9.9 1368 13 Q8UW85
33 659 9.7 1418 13 Q93457
34 640.5 9.5 1369 13 Q8UW86
35 634.5 9.4 1472 5 Q9U5A8
36 625 9.2 1412 13 Q8UW84
37 620 9.2 1358 13 Q73798
38 608.5 9.0 1418 13 Q8UW83
39 601 8.9 1245 13 Q9VGH8
40 588 8.7 1371 11 Q9QVW4
41 581.5 8.6 1091 4 Q9UMQ4
42 579 8.5 987 11 Q91YMO
43 578.5 8.5 2144 5 Q9VD94
44 576 8.5 935 4 Q96L35
45 576 8.5 987 11 Q99MR2

```

ALIGNMENTS

```

RESULT 1
O18735
ID O18735 PRELIMINARY; PRT; 1259 AA.
AC O18735;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Erbb-2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Pissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Yokota H.;
RT "cDNA cloning of erbB-2 from canine mammary gland.";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB008451; BAA23127.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Pfam; PF02757; YLP_2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1259 AA; 137989 MW; E37364D49C4ACD46 CRC64;

```

```
Query Match      89.9%; Score 6091; DB 6; Length 1259;
Best Local Similarity 89.5%; Pred. No. 0;
Matches 1128; Conservative 44; Mismatches 75; Indels 14; Gaps 4;

Qy 1 MELAAALCRWGLLLALLPFGAASTOCTGTDMKRLPASPETHLDMLRHLVGGCGVOGNL 60
Db 1 MELAAWCRWGLLLALLPFGAAGTOCTGTDMKRLPASPETHLDMLRHLVGGCGVOGNL 60

Qy 61 ELTVLPNASSLSFQDIQEOGVYVLIHAHNOVROVPLORLRIVRGTQLPEDNYALAVLDNG 120
Db 61 ELTVLPANASSLSFQDIQEOGVYVLIHAHQVRIPLQRLRIVRGTQLPEDNYALAVLDNG 120

Qy 121 DPLNNTTVPVTGASPGGLRELQRLSLTEILKGVVLIQRNPOLICYQDITLWKDIFHNKOLA 180
Db 121 DPLEGGIPAPGAAGGLRELQRLSLTEILKGVVLIQRNPOLCHQDITLWKDVHFHNKOLA 180

Qy 181 LTLIDTNRSRACHPCSPMKGSRGWSGSSDCOSLRTVTCAGGCARCKGPLPTDCCHQEC 240
Db 181 LTLIDTNRFRSACPPSPACKDAHCWAGSSGDCSLRTVTCAGGCARCKGQPTDCCHQEC 240

Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTSCP 300

Qy 301 YNYLSTDVGSCTLVCPLNHOEVTAEDGTQRCCKSKPCARVCYGLGMQYIKANSKFITGIT 360
Db 301 YNYLSTDVGSCTLVCPLNHOEVTAEDGTQRCCKSKPCARVCYGLGNEHLREVRAVTSAN 360

Qy 361 ELEPAGCKKIFGSLAFIPESFGDGPASNTAPLOPEQLQVFETLEETIGYLYISAWPDSLP 420
Db 361 IQEPAGCKKIFGSLAFIPESFGDGPASNTAPLOPEQLRVFEALEETIGYLYISAWPDSLP 420

Qy 421 DLSVFQNLQVTRGRLHNGAYSILTOGLGISWGLRSLRGLSGSLALIHNTLHLCFVHTV 480
Db 421 NLSVFQNLVIRGRLVHDGAYSILTOGLGISWGLRSLRGLSGSLALIHNRALCFVHTV 480

Qy 481 PWDOLFNPQHALLHTANRDECECVGEGGLACHQICARGHCWGPGPTQCVNCSQFLRQEC 540
Db 481 PWDOLFNPQHALLHSANRDEECVGEGLACYP-CAHGHGWCWGPGPTQCVNCSQFLRQEC 539

Qy 541 VEECRVLQGLPREVYNARHCLPCHPECPQNGSVTCFPGPADQCACAHYKDPFPCVARC 600
Db 540 VEECRVLQGLPREYVKDRYCLPCHSECQPNGSVTCFSGEADQCACAHYKDPFPCVARC 599

Qy 601 PSGVKPDLSTWMPYKWPDESGACQPCPINCTHSCVDLDDKGPABQASPLTSTVSAVG 660
Db 600 PSGVKPDLSTWMPYKWFADSEGTQPCPINCTHSCADLDKGPABQASPLTSLIAAVG 659

Qy 661 ILLVVLGVVFGIILIKRRQKIRKYTMRLLIQETELVEPLTPSGAMPNQAMRILKETEL 720
Db 660 ILLAVVGLVLGILI KRRQKIRKYTMRLLIQETELVEPLTPSGAMPNQAMRILKETEL 719

Qy 721 RKVKVLSGAFGVYKA-----NSKPIGITELVLRENTSPKANKEILDEAYVMAGVGS 773
Db 720 RKVKVLSGAFGVYKGIWIPDGENVK-IPVAIKVLRENTSPKANKEILDEAYVMAGVGS 778

Qy 774 PYVSRLAGICLTSVQLVLTQMPVGCCLLDVHRENRGLSGQDLLNWCQIAKGMVLEDV 833
Db 779 PYVSRLAGICLTSVQLVLTQMPVGCCLLDVHRENRGLSGQDLLNWCQIAKGMVLEDV 838

Qy 834 RLVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHYADGGKVPKMMALRESILRRRF 893
Db 839 RLVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHYADGGKVPKMMALRESIPPRRF 898

Qy 894 THOSDVMWSYGVVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVTVMWVKW 953
Db 899 THOSDVMWSYGVVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVTVMWVKW 958

Qy 954 MIDSECPRELVSEFSRWARDPORFVIONEDLGASPLDSTFYRESLEDDMDGLVD 1013
Db 959 MIDSECPRELVSEFSRWARDPORFVIONEDLGASPLDSTFYRESLEDDMDGLVD 1018

Qy 1014 AEEYLVPOQGFPCPEPTPGAGGTAAHRRHSSSTRNGGELTLGLEPSEEBPKSLAPSE 1073
Db 1019 AEEYLVPOQGFPCPEPTPGAGGTAAHRRHSSSTRNGGELTLGLEPSEEBPKSLAPSE 1078

Qy 1074 GAGSDVFDGDLGMAAKGLQSLTHDPSPLORYSEDPTVPLPSETDGYVAPLTCSPQPEY 1133
Db 1079 GAGSDVFDGDLGMAAKGLQSLTHDPSPLORYSEDPTVPLPSETDGYVAPLTCSPQPEY 1138

Qy 1134 VNOPDVRPOPSPSPREGPLPAARPAAGATLER-----AKTILSPGKNGVVKVDFATGGAVENP 1188
Db 1139 VNOPDVRPOPSPSPREGPLPAARPAAGATLERPKTILSPGKNGVVKVDFATGGAVENP 1198

Qy 1189 EYLTPQGAAPQPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPBYLGLDVP 1248
Db 1199 EYLTPQGAAPQPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPBYLGLDVP 1258

Qy 1249 V 1249
Db 1259 V 1259

RESULT 2
Q9QX70 PRELIMINARY; PRT; 1209 AA.
AC Q9QX70;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Epidermal growth factor receptor.
GN EGFR.
OS Rattus norvegicus (Rat).
OC Mammalia; Eutharia; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutharia; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FISHER; TISSUE=LIVER;
RX MEDLINE=90258888; PubMed=2342466;
RA Petch L.A., Harris J., Raymond V.W., Blaeband A.J., Lee D.C.,
RA Earp H.S.;
RT "A truncated, secreted form of the epidermal growth factor receptor is
RT encoded by an alternatively spliced transcript in normal rat tissue."
RL Mol. Cell. Biol. 10:2973-2982(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FISHER; TISSUE=LIVER;
RA Petch L.A.;
RL Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=FISHER; TISSUE=LIVER;
RA Guttridge K., Dawson T.L., Earp H.S.;
RL Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.
DR EMBL; M37394; AAF14008.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00111; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
KW ATP-binding; Receptor; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1209 AA; 134891 MW; 96FEE7F6CC1B7773 CRC64;

Query Match      44.7%; Score 3028; DB 11; Length 1209;
Best Local Similarity 49.1%; Pred. No. 8.6e-221;
```

	Matches	629;	Conservative	165;	Mismatches	360;	Indels	126;	Gaps	31;
Qy	3	LAALCRGLLLALLPPGA	-ASTOVC	TG	DMKRLRLPAS	ETHLDMRLRHLYQCQCVVQGNLE	61			
Db	15	LAALCAAG-----	-GALBEKKV	CGGT	SNRLTQ	GFDPEDHFLSLQRMFNNCCEVVLGNLE	66			
Qy	62	LTLYPTWASLSFLODIOE	VQGYVLI	AHNOVQ	RVPLRLRIVRG	TQLFDNVALAVLNDGD	121			
Db	67	ITVYVORNYDLISFLKTI	QEVAGYVLI	ALNTVBERI	PLENLQIIRGN	ALYENTVALAVLSN--	124			
Qy	122	PLNNTTPVTGASPGGLR	ELEQLSLTEILK	GGVLIQ	RNLPOLCVQ	DTILWKDIFHKNLQAL	181			
Db	125	-----	YGTNTK	TGLRELPMRLQ	EILIGAVR	SNPILCMETIQWRD	175			
Qy	182	TLIDTNRS-RACHPC	SPMKSGRCW	GSESSDCQ	SLRTVCAGGCA	-RCKGPLPTDCCHEQ	239			
Db	176	MSMDVQRHLTGCPK	CDPSCPN	SGCWGEENCQ	LTKIICAQ	CSRRCRGRSPSDCCHNQ	235			
Qy	240	CAAGCTGPKUSDDCL	ACHLHFN	ISGICELH	CPALVTYNTDT	FESMPNPEGRYTFGASCVTAC	299			
Db	236	CAAGCTGPRSDCLV	CHFRDEAT	CKDTCPP	MLYNTPTYQ	MDVNPGEKYSFGATCVKCK	295			
Qy	300	PYNVLS	TDVGSCTLV	CPHNOE	VTAEQGTOR	CEKCSKPCARVCYGLGMQYIKANSKFTGI	359			
Db	296	PRNTV	TDHGS	CVRACG	PDYIEV	-EEDGVSCKCKDGPCKVCNGIGIGEPK-DTLSINA	353			
Qy	360	TELE-FAGCKK	IFGSLAF	LPESFDQ	PASNTAP	LQPEQLQVFELEETIGLYLISAWPDS	418			
Db	354	TNIKHFYCYTAI	SGDLHIL	VAFKGS	SFTRPPLD	PRELEILKTVKEITGFLLLQAMPEN	413			
Qy	419	LPDLSFQNTQVIR	GRILHNG	AYSLLTQ	GLGISWGL	RSRLBSGLALIHNNHTLCFVH	478			
Db	414	WTDLHAFENLEI	IRGTRKHQ	QGSFLAV	VVLNITS	LGLRSLKEISDGDVITSGNRLCYAN	473			
Qy	479	TVPNDQLFRPHQ	ALLHNP	PEDE	CVGEG	LACHQLCARGHCWGPGTQCVCNCSQFLRGO	538			
Db	474	TINWKULFGT	PNQTKIK	IMNNRAE	KDKATN	HVCMPLCSGSCGCGPEPTDCVSCQNVSRGR	533			
Qy	539	ECVEEC	VLQCLPREY	NARHCL	PCHECQ	PQNGSVTCFGPEAQCVACAHYKDPFPFCA	598			
Db	534	ECVDKCNILEGE	PREFVENSE	CIOCHPE	CLPQTN	IICTGRGPNCTICAHYVDGPHCVK	593			
Qy	599	RCPSGVKPDLSYMP	IWKPP	BEGACQ	PCP	PNCTHSCVDLDDKGPAPQORASP-LTISVSA	657			
Db	594	TCPSGIMGENNTL	-VMKFAD	ANNVCHL	CHANCTY	GCGPGLKGC-QQPEGPKIPSIATG	650			
Qy	658	VVGTLLVVVLGV	FGI-LIK	EROQKIRKY	TWRLLO	ETELVEPLTPSGAMPNQAMRLK	716			
Db	651	IVGGLLFTIV	-VALGIGL	FMRRRLQ	VKRRTL	RRLLQRELVPELTPSGEAPNQAHRLTK	709			
Qy	717	ETELURKVKL	SGGA	FGO-----	YIKANSK	F--TGITELVRENTSPKANKEILDEAYVM	768			
Db	710	ETEPFKIKVL	SGA	FGTVYKGL	WTPEGEK	VIPIVAKE--UREATSPKANKEILDEAYVM	767			
Qy	769	AGVGS	PVYRLLGLCL	STVOL	TQ	LMPYGCGLD	VHNRNRLGSDLLNWCQIAKMS	828		
Db	768	ASVDNPHV	RLGLCL	STVQLITQ	LMPYGC	LLDVYREHKDNIGSQYLNNWCQVIAKGMN	827			
Qy	829	YLEDVRLVHRD	LAARNVL	VKSPNH	VKTDFGL	ARLLDIDETEHADGKQVPIKWMALESI	888			
Db	828	YLEDRLVLRD	LAARNVL	VKTPQ	HVKTDFGL	AKLLGAEEXEYHAEQGVPIKWMALESI	887			
Qy	889	LRRRFTHQ	SDVWSYGV	TWELMT	FGAKPYDGI	PARETPDILLEKGERLPOPPICTIDVYMI	948			
Db	888	LHRIYTHQ	SDVWSYGV	TWELMT	FGSKPYDGI	PASEISSILLEKGERLPOPPICTIDVYMI	947			
Qy	949	MVKCWMID	SECRPR	PRELVS	FSRMARD	PQRVVVIQ-NEDLGPASPLDSTFYRSLLEDD	1007			
Db	948	MVKCWMID	ADSRK	PREL	ILIFSR	MDARDPQRYLVIQGD	ERMHLPSDTSNFRALMEED	1007		
Qy	1008	MGDLVDAE	EYLVPOQ	GFCDP	DPAPAC	AGGVMHRRHSSTRSGGDLTLGLPEPSSEEPAPRS	1067			
Db	1008	MEDVVD	ADAEYLV	POQGF	-----	NSPST-----SRT	1033			


```
Qy 489 PHQALLTANRPEDECVGEGEGLACHQLCARGHCWGPGPTQCVCNCSQFLRGQCEVCECRVLQ 548
Db 450 SDHEVMVQKRNATECHEEGMECEQSCAGCKGKEQCEQCLCKNVKYKGCCLDSCK--- 506
Qy 549 GLPREY-VNARHCLPCHPECPQNGSVTCFPEADQCVACHYKDPDFVARCP----- 601
Db 507 SLPLRYSDVSKTCDGCHQCKD-----FCYPNEDNCGSCNVKDGRCFVACBCTTKHAM 561
Qy 602 -----SGVKPDLSTYMPIWKPD----- 618
Db 562 NGTCINCHKTCVCGRPRDTIAPGCTICSDKAIIGSDAKIERCLMKDESCPDGYSDYVL 621
Qy 619 -EEG----- 621
Db 622 QEEGLQKLSGKAVCRKCHPRCKKTGVGFHEQCECTGYKGEQCEDECPQDFYANEE 681
Qy 622 --ACQCPINT--HSCVDL-----DD-----KGPAAEQ----- 646
Db 682 TRICLPCHQECRGCHGLGDDHHCRLKLFEGDPYDNATFTCVSNCPASHPYKRFQEA 741
Qy 647 -----RASPLTISVAVGVGILLVVVLGVVFGI---LIKRRQKIRKVTM 687
Db 742 GKIGPYCSADSMQGLRIEPTQKIVNGSVNALILLCVFGIAFVLSRHNKKDAVKM 801
Qy 688 RLLQETELVEPLTFPGAMPNQAQRILKETELRKVKVLGSGFARQYIK-----ANSKF 741
Db 802 TMLAGCEDSPLRPSNVGNPLTKLRIKEAEIRRGVGLMGAFGRVFKGVWMEGESVK 861
Qy 742 IGITELVIRENTSPANKELIDRAYMAGVSPVSRLLGLCLTSTVOLVQLMPYGLL 801
Db 862 IPVAIKVLMEMSGSESSKEFELEAYINASVEHPNLLKLLAVCMTSQMWLITQLMPLGCLL 921
Qy 802 DHVENRGLGSODLLNWCIOIAGMSVLEDRVLVHRDLAARNLVKSPNHVKITDRGLA 861
Db 922 DYVRNKKDKISKALLNWSQIARGMAYLEBRRVLVHRDLAARNLVQTPSCVKITVGLA 981
Qy 862 RLLDIDETEHADGGKVPKIMMALESILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIP 921
Db 982 KLLDFDSEYRAAGGKMPKWLALCIRHRVFTSKSDVWAFGIIITWELLTYGARPENVP 1041
Qy 922 ARETPDLLEKERLPPOPICTIDYIMVMVKWMTDSECRPRFRELVSFESMARDPORFV 981
Db 1042 AKOYPELIEIGHKLPOPDICSLDYVICLLSCWILDADARPTFKQLAETFAEKARDPGRYL 1101
Qy 982 VIQNEGLGASPLDSTFVRSLEDDMDGLV----- 1012
Db 1102 MI-----PGDKFRLPSYTNQDEKDLIRTLAPVMAAAAAAAGASNVDPVPTIA 1152
Qy 1013 DAEYLVPOQGFCCPDPAAGAGMVMHRRSSSTRSGGDLTLGLEPSEEEAPRS----- 1067
Db 1153 ETDEYLQPKTRPSIMLPGPSA-----VEPS-DEMPKSLRYCK 1188
Qy 1068 -PLAP---SEGAGSDVFDGLGMAAGKGLQLPHTDPSPLORYSEDTVPPLPSSTDGYVA 1123
Db 1189 DPLKPDDETGDHGEV-----GVGGIR-----LNLPLDEDDYLM 1222
Qy 1124 PLTCSPOEYVQPDVVRPPSPREGPLPAARPAGATLERAKTILSPGKNGVVKDVFAFG 1183
Db 1223 P-TCQS-----NQS-----TPG-----YMDLIGVPA 1243
Qy 1184 AVENPEYL-----TPQGAAAPQPHPPAFSPAFDNLVYWDQDPPPERGAPSTFKGT 1234
Db 1244 SVDNPEYLMGSTQAIAGLQSGMG--PHTPP-----PNTFNGM 1280
Qy 1235 PTAENPE 1241
Db 1281 PTHQHSQ 1287

RESULT 8
Q9UK79
ID Q9UK79
AC Q9UK79; PRELIMINARY; PRT; 419 AA.
```

```
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Herstatin.
GN HER-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99415951; PubMed=10485918;
RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
RT autoinhibitor.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF177761; AAD56009.2;
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L domain; 1.
DR SMART; SMO0261; FU; 1.
SQ SEQUENCE 419 AA; 45472 MW; FEC1BE347E2D030C CRC64;

Query Match 27.6%; Score 1871; DB 4; Length 419;
Best Local Similarity 98.8%; Pred. No. 1.3e-133;
Matches 341; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MELAAALCRWGLLALLPPGAASQVCTGTDMLKRLPASPEHLDMLRHLHYGQGVVQGNL 60
Db 1 MELAAALCRWGLLALLPPGAASQVCTGTDMLKRLPASPEHLDMLRHLHYGQGVVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIQEVQGVVLIAHNOVQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIQEVQGVVLIAHNOVQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
Qy 121 DPLNTPPTVGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNOLA 180
Db 121 DPLNTPPTVGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNOLA 180
Qy 181 LTLIDNRSRACHPCSPCKSGSCWGESSEDCSLTRTVCCAGCARCKGLPTDCCHEQC 240
Db 181 LTLIDNRSRACHPCSPCKSGSCWGESSEDCSLTRTVCCAGCARCKGLPTDCCHEQC 240
Qy 241 AAGCTGPKHSDCLACLHFNHSGICEHLCPALVYNTDTFESMPNPEGRTYFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLHFNHSGICEHLCPALVYNTDTFESMPNPEGRTYFGASCVTACP 300
Qy 301 YNYLSTDVSGCTVLCPLHNOEVTAEQGTQRCCKSCPKARVCYGL 345
Db 301 YNYLSTDVSGCTVLCPLHNOEVTAEQGTQRCCKSCPKARVTHSL 345

RESULT 9
Q8R2X1
ID Q8R2X1 PRELIMINARY; PRT; 367 AA.
AC Q8R2X1;
DT 01-JUN-2002 (T-EMBLrel. 21, Created)
DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Hypothetical 40.2 kDa protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
```

```
DR EMBL; BC027080; AAH27080.1; -.
KW Hypothetical protein.
SQ SEQUENCE 367 AA; 40163 MW; 0BE03395F9E101B0 CRC64;

Query Match      25.7%; Score 1739; DB 11; Length 367;
Best Local Similarity 88.0%; Pred. No. 1.1e-123;
Matches 323; Conservative 11; Mismatches 33; Indels 0; Gaps 0;

Qy 883 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICT 942
Db 1 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICT 60

Qy 943 IDVTMIMVKWMDISCRPRFRELVSFESRMARDPQRFVVIQNEIDLGASPLDSTFYRSL 1002
Db 61 IDVTMIMVKWMDISCRPRFRELVSFESRMARDPQRFVVIQNEIDLGASPLDSTFYRSL 120

Qy 1003 LEDDDMGDLVDAEYLVPOQGFPCPDAPGAGGMVHRRSSSTRSGGDLTLGLEPSEE 1062
Db 121 LEDDDMGDLVDAEYLVPOQGFPCPDAPGAGGMVHRRSSSTRSGGDLTLGLEPSEE 180

Qy 1063 EAPRSPLAPSEGAGSDVDFDGLGMAAGKLGSLPHTDPSPLQRYSEDPTVPLPSETDGYV 1122
Db 181 EAPRSPLAPSEGAGSDVDFDGLGMAAGKLGSLPHTDPSPLQRYSEDPTVPLPSETDGYV 240

Qy 1123 APLTCSQPEYVNPQDVRPQPPSPREGPLPAARAGATLERAKTSLPGKNGVVKDVPAFG 1182
Db 241 APLTCSQPEYVNPQDVRPQPPSPREGPLPAARAGATLERAKTSLPGKNGVVKDVPAFG 300

Qy 1183 GAVENPEYLTPOGGAAPQP-----HPPA---ESPAPDNL 1214
Db 301 GAVENPEYLTPOGGAALSPDLLPSAQPSTTSITGTRTHQSGLHPAPSKGHLRQRTQST 360

Qy 1215 YVWD-QDPPER-----GAPPSTFKGTPTAEN 1239
Db 361 WWTQCEGQVRRSPDYSSGREGLTSAGIKWEGPPTTSRGTCARN 410

RESULT 10
Q8WYVO ID Q8WYVO PRELIMINARY; PRT; 412 AA.
AC Q8WYVO;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 44.7 kDa protein.
GN P31659.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
RA Wan D.F., Gu J.R.;
RT "Novel human cDNA clones with function of inhibiting cancer cell
RT growth."
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF118349; AAL55856.1; -.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
KW Hypothetical protein.
SQ SEQUENCE 412 AA; 44702 MW; 034397FF3F27D2BC CRC64;

Query Match      25.1%; Score 1697.5; DB 4; Length 412;
Best Local Similarity 80.5%; Pred. No. 1.9e-120;

Qy 569 PQNGSVTCFGEADQCACAHYKDPFCVARGCSGVKPLSYMPIWKFPDEGACQCPPI 628
Db 1 PQNGSVTCFGEADQCACAHYKDPFCVARGCSGVKPLSYMPIWKFPDEGACQCPPI 628

Matches 344; Conservative 75; Mismatches 124; Indels 112; Gaps 18;

Query Match      23.7%; Score 1604; DB 15; Length 729;
Best Local Similarity 52.5%; Pred. No. 5.5e-113;
Matches 344; Conservative 75; Mismatches 124; Indels 112; Gaps 18;

Qy 569 PQNGSVTCFGEADQCACAHYKDPFCVARGCSGVKPLSYMPIWKFPDEGACQCPPI 628
Db 1 PQNGSVTCFGEADQCACAHYKDPFCVARGCSGVKPLSYMPIWKFPDEGACQCPPI 628
```

```

Db 141 PEETATPKTGP--DHCMKCAFIDGPHCVKACPAAGVLGENDTL-VMKYADANAVCOLCHP 197
Qy 629 NCHTSCVDLDKGCAPAEQASPLTSIVSAV-GILLVVVLGVVFGILIKRQOKIRKVTM 687
Db 198 NCTGCKGPGLEGCP---NGSKTPSIAAGVVGGLCLVVGGLGIGLYLRRR-HIVRKRTL 253
Qy 688 RLLQETELVPLTPSGAMPNQAOMRILKETELKRVKVLGSGAFGQ-----YIKANSKF- 741
Db 254 RLLQERELVEPLTPSGEAPNQAHLRIKETEFKVKVLGSGAFGTVYKGLWIPGEKVK 313
Qy 742 --IGITELVRENTSPKANKEILDEAYVMAGVSPYVSRLIGICTSTVQLVTOLMPYGC 799
Db 314 IPVAIKS--LREATSPKANKEILDEAYVMASVDNPRVCRLLIGICTSTVQLITQMPYGC 371
Qy 800 LLDHVRENRGLSGODLLNWCQIAKMSYLEDLVLRDLAARNVLKSNHVKITDFG 859
Db 372 LLDVIREKHONIGSOYLLNWCQIAKGMNLEERLVRDLAARNVLKTPQHVKITDFG 431
Qy 860 LARLLDIDETEHADGGKVPKWMALLESILRRRTHQSDVMSYGVTVWELMTFGAKPYDG 919
Db 432 LAKLGADEKEYHAEGGKVPKWMALLESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDG 491
Qy 920 IPAREIDPLEKGRLLPOPPICITIDVYIMVKMWIDSECRPRFRELVSFSRWARDPQR 979
Db 492 IPASEISSVLEKGRLLPOPPICITIDVYIMVKMWIDADSRPKFRELIAEFSKWARDPPR 551
Qy 980 FVVIQ-NEDLGPASPLDSTFYRSLLDDMDGLVDAEYLVPOGFCPPDPAPGAGMVH 1038
Db 552 YLVIQGERMHLPSPTDSKFYRTLMBEEDMEDI VDAEYLVPHQGF----- 598
Qy 1039 HRHRSSTRSGGDLTLGLEPSEBEAPRSP-----APSEGAGSDVDFDGLMGAAKGLQ 1093
Db 599 ---NSPST-----SRTPLLSLSATSNSATNCID-----RNGQ 629
Qy 1094 SLPHDPSLPQRYSEDPTVPLPSET--DGYVAPLTCSPQPEYVNPQDVRPQPPSPREGPL 1151
Db 630 GHPVREDSFVQRYSSDPTGNFLEESIDDGFL-----PAPEYVNO--LMPKKPS----- 675
Qy 1152 PAARPAGATLERAKTLPSPKNGVVKDVF-----AFGGAVENPEYL 1191
Db 676 -----TAMVQIQIYNNISLTAISKLPMDSRVQNSHSTAVDNPEYL 715

RESULT 12
Q86714
ID Q86714 PRELIMINARY; PRT; 567 AA.
AC Q86714
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE V-erbB protein (Fragment).
GN V-ERBB.
OS Avian rous-associated virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60727.1; -.
DR HSSP; P11362; IFGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrcK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.

```

```

DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
KW TYROSINE; PS00109; PROTEIN_KINASE_TYR; 1.
FT NON_TER 1
SQ SEQUENCE 567 AA; 63390 MW; C6D9CBA7ADF725E1 CRC64;

Query Match 23.6%; Score 1602; DB 15; Length 567;
Best Local Similarity 53.1%; Pred. No. 5.3e-113;
Matches 343; Conservative 73; Mismatches 118; Indels 112; Gaps 18;

Qy 578 GPBDAQCVACHYKDPFCVAPCPGKVPKPLSVMYPIWKPFDEGACQCPCEINTHSCVDL 637
Db 1 GP--DHCMKCAFIDGPHCVKACPAAGVLGENDTL-VMKYADANAVCOLCHPNCRTGCKGP 57
Qy 638 DDKGCAEQASPLTSIVSAV-GILLVVVLGVVFGILIKRQOKIRKVTMRLQETEL 696
Db 58 GLEGCP---NGSKTPSIAAGVVGGLCLVVGGLGIGLYLRRR-HIVRKRTLRLLOEREL 113
Qy 697 VEPLTPSGAMPNQAOMRILKETELKRVKVLGSGAFGQ-----YIKANSKF---IGITELV 748
Db 114 VEPLTPSGEAPNQAHLRIKETEFKVKVLGSGAFGTVYKGLWIPGEKVKIPVAIKE-- 171
Qy 749 LRENTSPKANKEILDEAYVMAGVSPYVSRLIGICTSTVQLVTOLMPYGCCLDHYVNR 808
Db 172 LREATSPKANKEILDEAYVMASVDNPRVCRLLIGICTSTVQLITQMPYGCCLDYIREHK 231
Qy 809 GRLGSDLLNWCQIAKMSYLEDLVLRDLAARNVLKSNHVKITDFGLARLLDIDE 868
Db 232 DNTGSOYLLNWCQIAKGMNLEERLVRDLAARNVLKTPQHVKITDFGLAKLGADE 291
Qy 869 TEYHADGGKVPKWMALLESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIDPL 928
Db 292 KEYHAGGKVPKWMALLESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIPIASEISSV 351
Qy 929 LEKGERLPPOPPICITIDVYIMVKMWIDSECRPRFRELVSFSRWARDPQRFVVIQ-NED 987
Db 352 LEKGERLPPOPPICITIDVYIMVKMWIDADSRPKFRELIAEFSKWARDPPRYLVIQGER 411
Qy 988 LGPASPLDSTFYRSLLDDMDGLVDAEYLVPOGFCPPDPAPGAGMVHHRSSSTR 1047
Db 412 MHLPSPTDSKFYRTLMBEEDMEDI VDAEYLVPHQGF-----NSPST- 454
Qy 1048 SGGGDLTLGLEPSEBEAPRSP-----APSEGAGSDVDFDGLMGAAKGLQLSPLTHDPS 1102
Db 455 -----SRTPLLSLSATSNSATNCID-----RNGQGHVREDSF 489
Qy 1103 LQRYSEDPTVPLPSET--DGYVAPLTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAGAT 1160
Db 490 VQRYSSDPTGNFLEESIDDGFL-----PAPEYVNO--LMPKKPS----- 526
Qy 1161 LERAKTLPSPKNGVVKDVF-----AFGGAVENPEYL 1191
Db 527 -----TAMVQIQIYNNISLTAISKLPMDSRVQNSHSTAVDNPEYL 566

RESULT 13
Q64895
ID Q64895 PRELIMINARY; PRT; 962 AA.
AC Q64895;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Gag, v-erb-A, v-erb-B protein.
GN GAG, V-ERB-A, V-ERB-B.
OS Avian erythroblastosis virus.
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=11861;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90206603; PubMed=1969616;
RA Bruskina A., Jackson J., Bishop J.M., McCarley D.J., Schatzman R.C.;
RT "Six amino acids from the retroviral gene gag greatly enhance the
RT transforming potential of the oncogene v-erb-B.";

```

```

RL  Oncogene 5:15-24(1990).
CC  -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC  -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR  EMBL; X52209; CAA36459.1; -.
DR  EMBL; X52211; CAA36459.1; JOINED.
DR  HSSP; P10828; 2NLL.
DR  InterPro; IPR000719; Euk_pkinase.
DR  InterPro; IPR000536; Hormone_rec_lig.
DR  InterPro; IPR001723; Steroid_receptor.
DR  InterPro; IPR001245; Tyr_pkinase.
DR  InterPro; IPR001628; Znf_C4steroid.
DR  Pfam; PF00104; hormone_rec; 1.
DR  Pfam; PF00069; pkinase; 1.
DR  Pfam; PF00105; zf-C4; 1.
DR  PRINTS; PR00398; STRDHOMONER.
DR  PRINTS; PR00047; STROIDFINGER.
DR  ProDom; PD000001; Euk_pkinase; 1.
DR  ProDom; PD000035; Znf_C4steroid; 1.
DR  SMART; SM00430; HOL1; 1.
DR  SMART; SM00219; TyrKc; 1.
DR  SMART; SM00399; Znf_C4; 1.
DR  PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
DR  PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR  PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR  PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW  ATP-binding; DNA-binding; Nuclear protein; Receptor;
KW  Transcription regulation; Transferase; Tyrosine-protein kinase;
KW  Zinc-finger.
SQ  SEQUENCE 962 AA; 108320 MW; 3C5AED791E4E95CE CRC64;

Query Match      22.8%; Score 1543.5; DB 15; Length 962;
Best Local Similarity 49.6%; Pred. No. 3.3e-108;
Matches 343; Conservative 72; Mismatches 152; Indels 125; Gaps 19;

Qy  541 VEECRVLQGLPRE-YVVAR-HCLP-----CHPEQ 568
Db  354 IEKCOESYLLAFEHYINRYRGNHIFHWSKLMKLVADLMIGAYHASRFLHMKVECPTELS 413

Qy  569 PQNGSVTCFGEADQCVACAHYKDPFCVACRCPGVKPDLSYMPIMKFPDDEEGACQPCPI 628
Db  414 PQE-----VGP--DHCWKCAHFIDGPHCVKACPVGLGENDTL-VKYYADANAVCOLCHP 465

Qy  629 NCHTSCVDLDDKGCFAORASPLTSIVSAVV-GILLVVLGVVFGILIKRQOKIRKYTM 687
Db  466 NCTRCCKPGLEGCP--NGSKTPSIAAGVVGGLCLVVGIGLYLRRR-HIVKRTL 521

Qy  688 RLLOETELVEPLTPSGAMPNQAMRILKETELRKVKVGLSGARGQYIK-----ANSKF 741
Db  522 RLLQERELVEPLTPSGEAPNQAHRLILKETEFKKVKVGLGFGAGFTYVKGWIPEGEKVT 581

Qy  742 IGITELVIRENTSPKANKEILDEAYVMAGVGSPPVSRLLGICLTSTVQLTQMLPYGCLL 801
Db  582 IPVAIKELRENTSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLTQMLPYGCLL 641

Qy  802 DHVENRGRIGSQDLNMCQIAKMSYLEDLVRLDAAARNLVKSPNHRKIDFGLA 861
Db  642 DYIREHKNIGSQYLLNWCVCQIAKGMNLEERHMHVRLDAAARNLVKTPQHVKITDFGLA 701

Qy  862 RLDDIDETEHADGGKVPKMALESILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIP 921
Db  702 KQLGADEKEYHAEGGKVPKMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIP 761

Qy  922 ARETPDLLEKGERLPQPPICITDVMYMWKWMIDSECRPRFRELVSERFARMARDPQFV 981
Db  762 ASELSVLEKGERLPQPPICITDVMYMWKWMGADSRPKFRELIAEFKMAKDPDPRYL 821

Qy  982 VIQ-NEDLGPASPLDSTFYRSLLEDDDGLVDAAEYLVFQQGFPFCDDPAPFAGMGVHHR 1040
Db  822 VIQDERMHLPSPTDSKFYRLTMEEDMEDIVDAEYLVPHQGF-----866

Qy  1041 HRSSTSGGGDLTLGLEPSEEEAPRSPFAPSEGAGSDVFDGLGMGAAGLQSLPHTDP 1100
Db  867 -NSPST-----SRTEPLLSLSATSN-----NSATKCIDRNGGH-- 898

```

```

Qy  1101 SPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYVNPQDVRPQPSPREGFLPAARFAGAT 1160
Db  899 -----PVREDGFL-----PAPEYVNO--LMPKKPSTAMVQIYVNYISLT 936

Qy  1161 -LERAKTILSPGKNGVVDVFAFGGAVENPEYL 1191
Db  937 AISKLPMDSFYQN-----SHSTAVDNPEYL 961

RESULT 14
Q85468 PRELIMINARY; PRT; 545 AA.
ID Q85468
AC Q85468;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Avian Erythroblastosis virus (1834) v-erbB gene.
OS Avian erythroblastosis virus.
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
CX NCBI_TaxID=11861;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88217326; PubMed=2897102;
RA Scotting P., Vennstrom B., Jansen M., Graf T., Beug H., Haymann M.J.;
RT "Common site of mutation in the erbB gene of avian erythroblastosis
RT virus mutants that are temperature sensitive for transformation.";
RL Oncogene Res. 1:265-278(1987).
DR ENBL; X06943; CAA30024.1; -.
DR HSSP; P11362; IFGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 545 AA; 60899 MW; 140DCE8CCA0F8AF4 CRC64;

```

```

Query Match      22.7%; Score 1535; DB 15; Length 545;
Best Local Similarity 52.5%; Pred. No. 6.2e-108;
Matches 330; Conservative 69; Mismatches 131; Indels 98; Gaps 16;

Qy  578 GPEADQCVACAHYKDPFCVACRCPGVKPDLSYMPIMKFPDDEEGACQPCINCHTSCVDL 637
Db  1 GP--DHCWKCAHFIDGPHCVKACPVGLGENDTL-VKYYADANAVCOLCHPNCRTGCKGP 57

Qy  638 DDKGCPAEQORASPLTSIVSAVV-GILLVVLGVVFGILIKRQOKIRKYTMRLLOTEL 696
Db  58 GLEGCP--NGSKTPSIAAGVVGGLCLVVGIGLYLRRR-HIVKRTLRLLOTEL 113

Qy  697 VEPLTPSGAMPNQAMRILKETELRKVKVGLSGARGQYIK-----ANSKFIGITELVLR 750
Db  114 VEPLTPSGEAFNQAHRLILKETEFKKVKVGLGFGAGFTYVKGWIPEGEKVTIPVAIKELR 173

Qy  751 ENTSPKANKEILDEAYVMAGVGSPPVSRLLGICLTSTVQLTQMLPYGCLLDHVENRGR 810
Db  174 EATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLTQMLPYGCLLDYIREHKON 233

Qy  811 LGSQDLNMCQIAKMSYLEDLVRLDAAARNLVKSPNHRKIDFGLARLLDIDETE 870
Db  234 IGSQYLLNWCVCQIAKGMNLEERHMHVRLDAAARNLVKTPQHVKITDFGLAKQLGADEKE 293

Qy  871 YHADGGKVPKMALESILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPAREIPDLLE 930
Db  294 YHAEGGKVPKMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIPASEISSVLE 353

Qy  931 KGERLPQPPICITDVMYMWKWMIDSECRPRFRELVSERFARMARDPQFVVIQ-NEDLIG 989
Db  354 KGERLPQPPICITDVMYMWKWMGADSRPKFRELIAEFKMAKDPDPRYLVIQGDGRMH 413

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:41:54 ; Search time 36.7573 Seconds
(without alignments)
4527.811 Million cell updates/sec

Title: SEQ4-710-730-12

Perfect score: 6775

Sequence: 1 MELAALCRWGLLALLPPGA.....TFKGTPTAENPEYGLDVPV 1249

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_101002.*
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6608	97.5	1255	21	Human heregulin 2
2	6608	97.5	1255	22	Human tyrosine kin
3	6608	97.5	1255	22	HER2 transgene pla
4	6608	97.5	1255	23	Human HER2 (ErbB2)
5	6602	97.4	1255	17	HER-2/neu protein.
6	6602	97.4	1255	20	Human HER-2/neu on
7	6602	97.4	1255	21	Human HER-2/neu pr
8	6602	97.4	1255	21	Amino acid sequenc
9	6602	97.4	1255	22	Human HER-2/neu pr
10	6602	97.4	1255	22	HER2/neu amino aci

11	6602	97.4	1255	23	AAE24067	Human Her-2 protei
12	6602	97.4	1255	23	AAE20479	Human Her-2/neu pr
13	6602	97.4	1255	23	AAE51143	Human Her-2/neu on
14	6602	97.4	1255	23	AAU77114	Human Her-2/neu po
15	6559	96.8	1433	14	AAE39568	Sequence of c-erbB
16	6438	95.0	1223	23	AAU98923	Human breast cance
17	6285	92.8	1200	21	AAE21208	Human HER-2/neu pr
18	5815.5	85.8	1256	21	AAE21199	Rat HER-2/neu prot
19	5815.5	85.8	1256	23	AAE51144	Rat HER-2/neu onco
20	5788.5	85.4	1256	21	AAE21206	Mouse Her-2/neu pr
21	5788.5	85.4	1256	22	AAE28860	Amino acid sequenc
22	5788.5	85.4	1256	23	AAE51151	Mouse Her-2/neu on
23	4817	71.1	919	21	AAE21203	Human HER-2/neu fu
24	4817	71.1	919	23	AAE51148	Her-2/neu extracel
25	4067.5	60.0	920	23	AAE51152	Mouse Her-2/neu ex
26	4067.5	60.0	926	23	AAE51153	Mouse Her-2/neu ex
27	3701	54.6	712	21	AAE21204	Human HER-2/neu fu
28	3701	54.6	712	23	AAE51149	Her-2/neu extracel
29	3552	52.4	782	18	AAW19764	Her2-GM-CSF immuno
30	3550	52.4	653	21	AAE21200	Extracellular HER-
31	3550	52.4	653	23	AAE51145	Human ErbB2 oncopr
32	3512	51.8	645	22	AAE60408	Human ErbB2 extrac
33	3512	51.8	645	22	AAE61593	DC8scFv-erbB2EC fu
34	3447	50.9	951	21	AAE44393	Extracellular port
35	3344	49.4	624	11	AAE08222	Rat HER-2/neu prot
36	3053.5	45.1	654	21	AAE21205	Rat HER-2/neu onco
37	3053.5	45.1	654	23	AAE51150	Amino acid sequenc
38	3030	44.7	1210	21	AAE19259	Human EGF receptor
39	3030	44.7	1210	21	AAE50616	Human Her-1 protei
40	3030	44.7	1210	23	AAE23019	Human epidermal gr
41	3028	44.7	1210	22	AAE50768	Amino acid sequenc
42	3028	44.7	1210	22	AAE68420	Human epidermal gr
43	2989	44.1	1210	23	ABP51768	Human protein for
44	2958	43.7	583	23	AAE20483	Human protein for
45	2958	43.7	587	23	AAE20481	Human protein for

ALIGNMENTS

RESULT 1
AAE2620
ID AAY92620 standard; Protein; 1255 AA.
XX
AC AAY92620;
XX
DT 10-AUG-2000 (first entry)
XX
DE Human heregulin 2 (Her2).
XX
KW Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;
KW self-protein; cancer; breast cancer; prostate cancer;
KW cell-associated peptide antigen; foreign epitope.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 1..173
FT /label= N-terminal
FT /note= "mature polypeptide"
FT Region 5..25
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT Region 59..73
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT Region 103..117
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT Region 149..163
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT Domain 174..323

FT Region /label= Cysteine_rich_domain
 FT 210..224
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Region
 FT 250..264
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain
 FT 324..483
 FT /label= Ligand_binding_domain
 FT Region
 FT 325..339
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Region
 FT 369..383
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Region
 FT 465..479
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain
 FT 484..623
 FT /label= Cysteine_rich_domain
 FT Region
 FT 579..593
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain
 FT 624..654
 FT /label= Transmembrane_domain
 FT Region
 FT 632..652
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Region
 FT 653..667
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain
 FT 655..1010
 FT /label= Tyrosine_kinase_domain
 FT Region
 FT 661..675
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Region
 FT 695..709
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Region
 FT 710..730
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain
 FT 1011..1235
 FT /label= C-terminal_domain

WO200020027-A2.
 13-APR-2000.
 05-OCT-1999; 99WO-DK00525.
 05-OCT-1998; 98DK-0001261.
 20-OCT-1998; 98US-0105011.
 (MEEBI-) M & E BIOTECH AS.
 Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I;
 Gautam A, Birk P, Karlsson G;
 WPI; 2000-349917/30.
 DR N-PSDB; AAA09455.
 Inducing immune responses to weakly immunogenic, tumor associated
 peptide antigens for the treatment of breast and prostate cancer
 Claim 62; Page 193-198; 220pp; English.
 This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of
 Her2 can be used in the claimed method as an autovaccine to induce a CTL
 response. Subdominant CTL epitopes, antibody binding regions and
 cysteine residues involved in disulfide bonds are preserved in the
 immunogenized forms. Regions suitable for the insertion of foreign T
 helper epitopes were identified (see features table). The method

CC is used for inducing immune responses against weakly immunogenic
 CC cell-associated peptide antigens (PA) such as those associated with
 CC cancers (self-proteins), e.g. human prostate specific membrane antigen
 CC (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).
 CC The method comprises effecting simultaneous presentation by antigen
 CC producing cells (APCs) of the animals immune system of: (1) at least 1
 CC CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1
 CC B-cell group derived from the cell-associated PA; and (2) at least 1
 CC first T helper cell group which is foreign to the animal. Analogues of
 CC human PSM, human Her2 and human/murine FGF8b comprising a substantial
 CC part of all known and predicted CTL and B-cell epitopes of the respective
 CC PA and including at least one foreign T helper epitope are also claimed.
 CC The method is used to treat prostate, prostate/breast or breast cancer
 CC when the PA is human PSM, FGF8b and Her2, respectively.
 XX
 SQ Sequence 1255 AA;
 Query Match 97.5%; Score 6608; DB 21; Length 1255;
 Best Local Similarity 97.4%; Pred. No. 0;
 Matches 1223; Conservative 5; Mismatches 20; Indels 8; Gaps 2;
 Qy 1 MELAALCRWGLLALLPPGAASTQVCTGDMKRLPASPETHLDMLRLHLYQGCVVQGNL 60
 Db 1 MELAALCRWGLLALLPPGAASTQVCTGDMKRLPASPETHLDMLRLHLYQGCVVQGNL 60
 Qy 61 ELTYLPTNASLSFLQDIOEVQGVLIAHNOVROVPLQRLRIVRGTOLEFEDNYALAVLDNG 120
 Db 61 ELTYLPTNASLSFLQDIOEVQGVLIAHNOVROVPLQRLRIVRGTOLEFEDNYALAVLDNG 120
 Qy 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILMKNDIFHKNQLA 180
 Db 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILMKNDIFHKNQLA 180
 Qy 181 LTLIDNRSRACHPCSPMKSGRCWGESSEDCOSLRTTVTCAGGCARCKGLPTDCCHEQC 240
 Db 181 LTLIDNRSRACHPCSPMKSGRCWGESSEDCOSLRTTVTCAGGCARCKGLPTDCCHEQC 240
 Qy 241 AACTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 Db 241 AACTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 Qy 301 YNYLSTDVSGCTLVCPLNQVETAEADGTQRCCKSPCARVCYGLGMQYIKANSKFIT 360
 Db 301 YNYLSTDVSGCTLVCPLNQVETAEADGTQRCCKSPCARVCYGLGMQYIKANSKFIT 360
 Qy 361 ELEFAGCKIFGSLAFIPESFDGDPASNTAPLOEQLOVPETLEEITGYLIYISAWPDSL 420
 Db 361 IQEFAGCKIFGSLAFIPESFDGDPASNTAPLOEQLOVPETLEEITGYLIYISAWPDSL 420
 Qy 421 DLSVFONLQVIRGRIHNGAYSILTLQGLISWLGSLRLSRELGSGLALIHNTLHLCFVHTV 480
 Db 421 DLSVFONLQVIRGRIHNGAYSILTLQGLISWLGSLRLSRELGSGLALIHNTLHLCFVHTV 480
 Qy 481 PWDQLFRNPQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGTQCVNCSQFLRGQEC 540
 Db 481 PWDQLFRNPQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGTQCVNCSQFLRGQEC 540
 Qy 541 VEESRVLOGLPREYVNAHCLPCHPECOPONGSVTCFGEADOCVACAHYKDPFPCVARC 600
 Db 541 VEESRVLOGLPREYVNAHCLPCHPECOPONGSVTCFGEADOCVACAHYKDPFPCVARC 600
 Qy 601 PSQVKPDLSPYMPIWKPDPDEGACQPCINCTHSCVDLDDKCPAEORASPLTISVSAVVG 660
 Db 601 PSQVKPDLSPYMPIWKPDPDEGACQPCINCTHSCVDLDDKCPAEORASPLTISVSAVVG 660
 Qy 661 ILLVVLGVVFGILIKRRQOKIRKYTMRLLOLQETELVEPLTPSGAMPNQAMRILKETEL 720
 Db 661 ILLVVLGVVFGILIKRRQOKIRKYTMRLLOLQETELVEPLTPSGAMPNQAMRILKETEL 720
 Qy 721 RKVKVLGSGAFQGYIYA-----NSKFTGITELVLRNTSPKANKEILDEATVMAGVGS 773
 Db 721 RKVKVLGSGAFQGYIYKGIWIPDGENYK-IPVAIKVLRNTSPKANKEILDEATVMAGVGS 779

```
Qy 774 PYVSRLLGICLTSTVQLTQMPYGCILLDHVRENRGRIGSQDILLNWCQIAKGMYSILEDV 833
Db 780 PYVSRLLGICLTSTVQLTQMPYGCILLDHVRENRGRIGSQDILLNWCQIAKGMYSILEDV 839
Qy 834 RLVRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKKNWALESLIRRRF 893
Db 840 RLVRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKKNWALESLIRRRF 899
Qy 894 THQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPPOPICTIDVYIMVKCW 953
Db 900 THQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPPOPICTIDVYIMVKCW 959
Qy 954 MIDSECPRRFELVSEFSRMARDQRFVWIONEDLGPASPLDSTFYRSLLDEDDDMGLVD 1013
Db 960 MIDSECPRRFELVSEFSRMARDQRFVWIONEDLGPASPLDSTFYRSLLDEDDDMGLVD 1019
Qy 1014 ABEYLVQOQFFCPDPAPGAGMVHRRHSSTRSGGDLTLGLEPSEEEAPRSPAPSE 1073
Db 1020 ABEYLVQOQFFCPDPAPGAGMVHRRHSSTRSGGDLTLGLEPSEEEAPRSPAPSE 1079
Qy 1074 GAGSDVDFDGLGMAAGKLSLTHDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEY 1133
Db 1080 GAGSDVDFDGLGMAAGKLSLTHDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEY 1139
Qy 1134 VNQPDVRRQPPSPREGPLPAARPAAGATLERAKTUSPGKNVGVKDVFAFGGAVENPEYLT 1193
Db 1140 VNQPDVRRQPPSPREGPLPAARPAAGATLERAKTUSPGKNVGVKDVFAFGGAVENPEYLT 1199
Qy 1194 QGGAAPQHPHPPAFSPAFDNLVYWDQPPERGAPPSTFKGTPTAENPEYLGLDVVP 1249
Db 1200 QGGAAPQHPHPPAFSPAFDNLVYWDQPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 2
AAE12130
ID AAE12130 standard; Protein; 1255 AA.
XX
AC AAE12130;
XX
DT 18-DEC-2001 (first entry)
XX
DE Human tyrosine kinase-type receptor, HER-2.
XX
KW Therapeutic compound; major histocompatibility complex; vaccine;
KW antigenic peptide; MHC; immunoregulatory; immune response; HER-2;
KW adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;
KW antigen presenting cell; human; tyrosine kinase-type receptor.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Region 774..782
FT /note="Antigenic epitope"
XX
PN WO200168677-A2.
XX
PD 20-SEP-2001.
XX
PF 16-MAR-2001; 2001WO-US40328.
XX
PR 16-MAR-2000; 2000US-0527487.
XX
PA (GENZ ) GENZYME CORP.
XX
PI Nicolette CA;
XX
WPI; 2001-616284/71.
DR N-ESDB; RAD19731.
XX
PT Novel synthetic therapeutic compound for inducing immune response and
PT for use in adoptive immunotherapy, has enhanced binding to major
PT histocompatibility molecules and enhanced immunoregulatory properties
```

XX

Claim 4; Page 63-67; 69pp; English.

The invention relates to synthetic therapeutic compounds (antigenic peptides) with enhanced binding to major histocompatibility complex (MHC) molecules and enhanced immunoregulatory properties relative to their natural counterparts. Compounds of the invention are useful for inducing an immune response in a subject and for use in adoptive immunotherapy. They are useful as components of anti-cancer vaccines and to expand immune effector cells that are specific for cancers characterised by expression of the breast cancer antigen, HER-2. Polynucleotides that encode peptides of the invention are useful as hybridisation probes and as primers for the detection of genes of gene transcripts that are expressed in antigen presenting cells (APCs), to confirm transduction of polynucleotides into host cells. The present sequence is human tyrosine kinase-type receptor, HER-2. Compounds of the invention are designed based on the HER-2 antigenic peptide (774-782).

SQ Sequence 1255 AA;

Query Match 97.5%; Score 6608; DB 22; Length 1255;
Best Local Similarity 97.4%; Pred. No. 0;
Matches 1223; Conservative 5; Mismatches 20; Indels 8; Gaps 2;

Qy 1 MELAALCRWGLLLALLPPGAASTQVCTGDMKRLPASPEHDLMLRHLHYQGCVVQGNL 60

Db 1 MELAALCRWGLLLALLPPGAASTQVCTGDMKRLPASPEHDLMLRHLHYQGCVVQGNL 60

Qy 61 ELTYLPTNASLSFLQDIOEVQGVYLIAHNQVROVPLORLIRIVRGTLFEDNYALAVLDNG 120

Db 61 ELTYLPTNASLSFLQDIOEVQGVYLIAHNQVROVPLORLIRIVRGTLFEDNYALAVLDNG 120

Qy 121 DPLNNTPTVTGASPGGLRELQLRSLTEILKGGVLIORNPOLCYQDTILWKDI FHKNNOLA 180

Db 121 DPLNNTPTVTGASPGGLRELQLRSLTEILKGGVLIORNPOLCYQDTILWKDI FHKNNOLA 180

Qy 181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVCCAGCARCKGLPTDCCHEQC 240

Db 181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVCCAGCARCKGLPTDCCHEQC 240

Qy 241 AAGCTGPKHSIDCLACILFHNHSGICEILHCPALVYNTDFTESMPNPEGRYTFGASCVTACP 300

Db 241 AAGCTGPKHSIDCLACILFHNHSGICEILHCPALVYNTDFTESMPNPEGRYTFGASCVTACP 300

Qy 301 YNYLSTDVSGSTLVCPILHNOEVTAEADGTORCEKSKPCARVCYGLGMOYIKANSKFEGIT 360

Db 301 YNYLSTDVSGSTLVCPILHNOEVTAEADGTORCEKSKPCARVCYGLGMOYIKANSKFEGIT 360

Qy 361 ELEFAGCKKIFGSLAFIPESFDGDPASNTAPLOEQLOVFPETLEEITGYLYISAWPDSLP 420

Db 361 IQEFAGCKKIFGSLAFIPESFDGDPASNTAPLOEQLOVFPETLEEITGYLYISAWPDSLP 420

Qy 421 DLSVFONQVIRORILHNGAYSILTLOGLISWLGSLRSLRELGSGLALIHNTLHCFVHTV 480

Db 421 DLSVFONQVIRORILHNGAYSILTLOGLISWLGSLRSLRELGSGLALIHNTLHCFVHTV 480

Qy 481 PWDOLFNRPHCALHTANRPEDECVGEGGLACHOLCARGHCWGPGTQCVNCSQFLRGQEC 540

Db 481 PWDOLFNRPHCALHTANRPEDECVGEGGLACHOLCARGHCWGPGTQCVNCSQFLRGQEC 540

Qy 541 VEESCRVLQGLPREYVNNARHCLPCHPBCQONGSVTCFGEADOCVACAHAHKDPFVCVARC 600

Db 541 VEESCRVLQGLPREYVNNARHCLPCHPBCQONGSVTCFGEADOCVACAHAHKDPFVCVARC 600

Qy 601 PSGVKPDLSYMPTWKFPDEBEGACQPCINCTHSCVDLDDKGCAPORASPLTSIVSAVVG 660

Db 601 PSGVKPDLSYMPTWKFPDEBEGACQPCINCTHSCVDLDDKGCAPORASPLTSIVSAVVG 660

Qy 661 ILLVAVLVGVVFGILIKRROOKIRKYTMRRLLQETELVEPLTPSGAMPNOAQMILKETEL 720

Db 661 ILLVAVLVGVVFGILIKRROOKIRKYTMRRLLQETELVEPLTPSGAMPNOAQMILKETEL 720

```
Qy 721 RKVKVLSGAFGQYIKA-----NSKFIGITELVRENTSPKANKEILD EAYVMAGVGS 773
Db 721 RKVKVLSGAFGTVYKGIWIPDGENVK-IPVAIKVLRNTSPKANKEILD EAYVMAGVGS 779
Qy 774 PYVSRLLGICLTSTVQLVTQMPYGCCLLDHVRENRLGSDLLNWCQIAKMSYLEDV 833
Db 780 PYVSRLLGICLTSTVQLVTQMPYGCCLLDHVRENRLGSDLLNWCQIAKMSYLEDV 839
Qy 834 RLVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMMALSIILRRF 893
Db 840 RLVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMMALSIILRRF 899
Qy 894 THQSDVMSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPPOPICTIDVYIMVKW 953
Db 900 THQSDVMSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPPOPICTIDVYIMVKW 959
Qy 954 MIDSECRPRFRELVSERSMARDQRFVVIQNEDLGPASPLDSTFYRSLLED DDMGDLVD 1013
Db 960 MIDSECRPRFRELVSERSMARDQRFVVIQNEDLGPASPLDSTFYRSLLED DDMGDLVD 1019
Qy 1014 AEEYLVPOQGFCDPAPAGAGMWHRRHRSSTSRSGGDLTLGLEPSEEAAPRSLAPSE 1073
Db 1020 AEEYLVPOQGFCDPAPAGAGMWHRRHRSSTSRSGGDLTLGLEPSEEAAPRSLAPSE 1079
Qy 1074 GAGSDVFDGDLGMAAGLQSLPTHDPSPLOQYSEDPTVPLPSETDGYVAPLTCSPQPEY 1133
Db 1080 GAGSDVFDGDLGMAAGLQSLPTHDPSPLOQYSEDPTVPLPSETDGYVAPLTCSPQPEY 1139
Qy 1134 VNQPDVRPQPPSPREGPLUPAARPAAGATLERAKTILSPGKNGVVKDVFAPGGAVENPEYLT 1193
Db 1140 VNQPDVRPQPPSPREGPLUPAARPAAGATLERAKTILSPGKNGVVKDVFAPGGAVENPEYLT 1199
Qy 1194 QGGAAPQHPHPPAPSPAFDNLVYWDQDPPPERGAPPSTFKGPTAENPEYLGDLV 1249
Db 1200 QGGAAPQHPHPPAPSPAFDNLVYWDQDPPPERGAPPSTFKGPTAENPEYLGDLV 1255

RESULT 3
AAB60167
ID AAB60167 standard; Protein; 1255 AA.
XX AC AAB60167;
XX DT 03-APR-2001 (first entry)
XX DE HER2 transgene plasmid construct encoded protein.
XX KW Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
XX KW antibody.
XX OS Homo sapiens.
XX OS Synthetic.
XX FN WO200100244-A2.
XX PD 04-JAN-2001.
XX PF 23-JUN-2000; 2000WO-US17229.
XX PR 25-JUN-1999; 99US-0141316.
XX PR 16-MAR-2000; 2000US-0189844.
XX PA (GETH ) GENENTECH INC.
XX FI Erickson S, Schwall R;
XX DR WPI; 2001-061962/07.
XX DR N-PSDB; AAF24297.
XX PT Treating tumors, particularly breast cancers, which overexpress an ErbB
XX PT receptor and does not respond to an anti-ErbB antibody, comprises
XX PT conjugating the antibody to a maytansinoid -
```

PS Example 3; Fig 4; 92pp; English.

XX CC The present invention provides a method of treating cancer by
XX CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In
XX CC particular, the antibody is directed against ErbB2 (also known as HER2
XX CC and p185neu). The method is particularly useful in the treatment of
XX CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
XX CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.

SQ Sequence 1255 AA;

Query Match 97.5%; Score 6608; DB 22; Length 1255;
Best Local Similarity 97.4%; Pred. No. 0;
Matches 1223; Conservative 5; Mismatches 20; Indels 8; Gaps 2;

```
Qy 1 MELAALCRWGLLALLPPGNAASTQVCTGDMKRLRPASPETHLDMLRHLHQGQVVOGNL 60
Db 1 MELAALCRWGLLALLPPGNAASTQVCTGDMKRLRPASPETHLDMLRHLHQGQVVOGNL 60
Qy 61 ELYLPTNASLSFLQDIEQVGVVLIHNVORVPLQRLIRVGTOLFEDNYALAVLDNG 120
Db 61 ELYLPTNASLSFLQDIEQVGVVLIHNVORVPLQRLIRVGTOLFEDNYALAVLDNG 120
Qy 121 DPLNNTTPTVGTASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDI FHKNNOLA 180
Db 121 DPLNNTTPTVGTASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDI FHKNNOLA 180
Qy 181 LTLIDNRSBACHPCSPMKSGSCWGESSEDCOSLRTVTCAGGACRCKGLPTDCCHEQC 240
Db 181 LTLIDNRSBACHPCSPMKSGSCWGESSEDCOSLRTVTCAGGACRCKGLPTDCCHEQC 240
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Qy 301 YNYLSTDVSGCTIVCPLEHNOEVTAEADGTQRCCKSPCARVCYGLGQYIKANSKFIT 360
Db 301 YNYLSTDVSGCTIVCPLEHNOEVTAEADGTQRCCKSPCARVCYGLGMEHREAVTSAN 360
Qy 361 ELEFAGCKKIFGSLAFIPESFDGDPASNTAPLOEQLOVFEETLEETGYLIYI SAWPDSLP 420
Db 361 IQEFAGCKKIFGSLAFIPESFDGDPASNTAPLOEQLOVFEETLEETGYLIYI SAWPDSLP 420
Qy 421 DLSVFQNLQVIRGRIHNGAYSILTLOGLSISWLSRLSRLSGLALIHNTLHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRIHNGAYSILTLOGLSISWLSRLSRLSGLALIHNTLHLCFVHTV 480
Qy 481 PWDQLFRNPHOALLHTANRPEDECVGEGLAHQLCARGHCWGPGPTQCVNCSQFLRGOEC 540
Db 481 PWDQLFRNPHOALLHTANRPEDECVGEGLAHQLCARGHCWGPGPTQCVNCSQFLRGOEC 540
Qy 541 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFEGPADQCVACAHYKDPFVCVARC 600
Db 541 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFEGPADQCVACAHYKDPFVCVARC 600
Qy 601 PSGVKPDLSTYMPKPPDEBACQPCPINCTHSCVDLDDKGCAPAEORASPLTSIVSAVVG 660
Db 601 PSGVKPDLSTYMPKPPDEBACQPCPINCTHSCVDLDDKGCAPAEORASPLTSIVSAVVG 660
Qy 661 ILLVWVGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
Db 661 ILLVWVGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
Qy 721 RKVKVLSGAFGQYIKA-----NSKFIGITELVRENTSPKANKEILD EAYVMAGVGS 773
Db 721 RKVKVLSGAFGTVYKGIWIPDGENVK-IPVAIKVLRNTSPKANKEILD EAYVMAGVGS 779
Qy 774 PYVSRLLGICLTSTVQLVTQMPYGCCLLDHVRENRLGSDLLNWCQIAKMSYLEDV 833
Db 780 PYVSRLLGICLTSTVQLVTQMPYGCCLLDHVRENRLGSDLLNWCQIAKMSYLEDV 839
Qy 834 RLVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMMALSIILRRF 893
```

Db 840 RLVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKWMALESILRRF 899
 QY 894 THQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPPOPICTIDVYIMVKW 953
 Db 900 THQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPPOPICTIDVYIMVKW 959
 QY 954 MIDSECPRELVSESRMARDQRFVVIQNEDLGPASPLDSTFYRSLLDDDDMGDLVD 1013
 Db 960 MIDSECPRELVSESRMARDQRFVVIQNEDLGPASPLDSTFYRSLLDDDDMGDLVD 1019
 QY 1014 ABEYLVPQGGFFCPDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEERAPRPLAPSE 1073
 Db 1020 ABEYLVPQGGFFCPDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEERAPRPLAPSE 1079
 QY 1074 GAGSDVFDGLGMAAGKGLSLPHTDPSPLQRYSEDFTVPLPSETDGYVAPLTCSPQPEY 1133
 Db 1080 GAGSDVFDGLGMAAGKGLSLPHTDPSPLQRYSEDFTVPLPSETDGYVAPLTCSPQPEY 1139
 QY 1134 VNQPDVTPQPPSPREGPLPAARPAAGATLRAKTLSPGKNGVVDVPAFGGAVENPEYLT 1193
 Db 1140 VNQPDVTPQPPSPREGPLPAARPAAGATLRAKTLSPGKNGVVDVPAFGGAVENPEYLT 1199
 QY 1194 QGGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGDLV 1249
 Db 1200 QGGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGDLV 1255

RESULT 4
 AAU74545
 ID AAU74545 standard; Protein; 1255 AA.
 XX AAU74545;
 DT
 XX
 XX
 DE Human HER2 (ErbB2) polypeptide.
 XX
 KW Human; HER2; ErbB; epidermal growth factor receptor; receptor;
 KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
 KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
 KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
 KW glial disorder; astrocytal disorder; hypothalamic disorder;
 KW glandular disorder; macrophagal disorder; epithelial disorder;
 KW stromal disorder; blastocoealic disorder; inflammatory disorder;
 KW angiogenic disorder; immunological disorder.
 XX
 OS Homo sapiens.
 XX
 PN US2002001587-A1.
 XX
 PD 03-JAN-2002.
 XX
 PF 16-MAR-2001; 2001US-0811123.
 XX
 PR 16-MAR-2000; 2000US-189844P.
 PR 05-OCT-2000; 2000US-238327P.
 XX
 XX (ERIC/) ERICKSON S.
 PA (SCHW/) SCHWALL R.
 PA (SLIW/) SLIWOWSKI M.
 XX
 XX Erickson S, Schwall R, Sliwowski M;
 XX WPI; 2002-163686/21.
 XX
 DR N-PSDB; ABK14058.
 DR
 XX
 PT Treating tumour characterised by overexpression of epidermal growth
 PT factor receptor, ErbB or cancer in mammal, comprises administering
 PT anti-ErbB antibody-maytansinoid conjugate to the mammal -
 XX
 XX Example 3; Fig 7; 93pp; English.
 PS
 XX The invention relates to treating a tumour in a mammal, where the tumour

CC is characterised by the overexpression of an epidermal growth factor
 CC receptor (ErbB) and does not respond or responds poorly, to treatment
 CC with an anti-ErbB antibody, comprising administering to the mammal an
 CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for
 CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
 CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
 CC prostate and bladder, preferably breast cancer. The breast cancer is a
 CC metastatic breast cancer or an aggressive form of metastatic breast
 CC cancer which overexpresses ErbB2. The method is also useful for treating
 CC neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal,
 CC epithelial, stromal, blastocoealic, inflammatory, angiogenic and
 CC immunological disorders. This sequence represents the human HER2 (ErbB2)
 CC polypeptide of the invention.

XX
 SQ Sequence 1255 AA;

Query Match 97.5%; Score 6608; DB 23; Length 1255;

Best Local Similarity 97.4%; Pred. No. 0;

Matches 1223; Conservative 5; Mismatches 20; Indels 8; Gaps 2;

QY 1 MELAALCRWGLLLALLPPGAASSTQVCTGDMKRLRLPASPEHLMLRLHLYQGQVVOG 60
 Db 1 MELAALCRWGLLLALLPPGAASSTQVCTGDMKRLRLPASPEHLMLRLHLYQGQVVOG 60
 QY 61 ELYLPTNASLSFLQDIOEVQGVVLIHQNQVROVPLQRLIRVRGTOLFDNYALAVLDNG 120
 Db 61 ELYLPTNASLSFLQDIOEVQGVVLIHQNQVROVPLQRLIRVRGTOLFDNYALAVLDNG 120
 QY 121 DPLNNTTPTVGTASPGGLRELQLRLTEILKGGVLIQORNPOLCYODTILWKDIFHNKOLA 180
 Db 121 DPLNNTTPTVGTASPGGLRELQLRLTEILKGGVLIQORNPOLCYODTILWKDIFHNKOLA 180
 QY 181 LTLIDNTRSRACHPCSPMKGSRGWESSEDCOSLTRTVCAGGCARCKGLPTDDCCHEQC 240
 Db 181 LTLIDNTRSRACHPCSPMKGSRGWESSEDCOSLTRTVCAGGCARCKGLPTDDCCHEQC 240
 QY 241 AAGCTGPKGSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 Db 241 AAGCTGPKGSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 QY 301 YNYLSTDVSGCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMYIKANSKF 360
 Db 301 YNYLSTDVSGCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMYIKANSKF 360
 QY 361 ELEFAGCKKIFGSLAFPLPESFDDGPASNTAPLOPEQLQVFPETLEEITGYLYISAWP 420
 Db 361 IQEFAGCKKIFGSLAFPLPESFDDGPASNTAPLOPEQLQVFPETLEEITGYLYISAWP 420
 QY 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGSLWGLSLRELGLALIHNTLHLCFVHTV 480
 Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGSLWGLSLRELGLALIHNTLHLCFVHTV 480
 QY 481 PWDQLFRNPHQALLHTANRPEDECVCEGLACHQLCARGHCWGPGTQCVCNSQFLRGQEC 540
 Db 481 PWDQLFRNPHQALLHTANRPEDECVCEGLACHQLCARGHCWGPGTQCVCNSQFLRGQEC 540
 QY 541 VEECRVQLGLPREYVNAHCLPCHPECPQONGSVTCFGEADOCVCAHYKDPFVCV 600
 Db 541 VEECRVQLGLPREYVNAHCLPCHPECPQONGSVTCFGEADOCVCAHYKDPFVCV 600
 QY 601 PSGVKPDLSPYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCFAEQRASPLTSIVSA 660
 Db 601 PSGVKPDLSPYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCFAEQRASPLTSIVSA 660
 QY 661 ILLVVLGVVFGILIKRQOKIRKYTMRLLOBTTELVEPLTPSGAMPNOAMRILKETEL 720
 Db 661 ILLVVLGVVFGILIKRQOKIRKYTMRLLOBTTELVEPLTPSGAMPNOAMRILKETEL 720
 QY 721 RKVKVLGSGAFGQYIKA-----NSKFIGITELVRENTSPKANKEIIDEA VVMAGV 773
 Db 721 RKVKVLGSGAFGQYIKGIWIPDGENVK-IPVAIKVIRENTSPKANKEIIDEA VVMAGV 779
 QY 774 PYVSRLLGICLTSTVOLVTQLMPYGLLDHVRNRRGLSGQDILLNWCMTAKGMSYLEDV 833

Db 780 PYVSRLLGICLTSTVQLVTPMPYGCILLDVRNKRGLGSDLLNWCWQIAKMSYLEDV 839
Qy 834 RLVRDLAARNVLYKSPNHVKITDFGLARLLDIDETEHADGGKVPKMALESILRRRF 893
Db 840 RLVRDLAARNVLYKSPNHVKITDFGLARLLDIDETEHADGGKVPKMALESILRRRF 899
Qy 894 THQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPOLLEKGERLPPOPICTIDVYMWKWCW 953
Db 900 THQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPOLLEKGERLPPOPICTIDVYMWKWCW 959
Qy 954 MIDSECRPRFRELVSFMRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVD 1013
Db 960 MIDSECRPRFRELVSFMRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVD 1019
Qy 1014 ABEYLVPQGFCDPAPGAGGMVHRHRSSTSRSGGDLTLGLEPSEEEAPRSLAPSE 1073
Db 1020 ABEYLVPQGFCDPAPGAGGMVHRHRSSTSRSGGDLTLGLEPSEEEAPRSLAPSE 1079
Qy 1074 GAGSDVFDGLGMAAGLQSLPHTDPSLQRYSEDPTVPLPSETDGVVAPLTCSPOPEY 1133
Db 1080 GAGSDVFDGLGMAAGLQSLPHTDPSLQRYSEDPTVPLPSETDGVVAPLTCSPOPEY 1139
Qy 1134 VNQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVYKDVFAFGGAVENPEYLTP 1193
Db 1140 VNQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVYKDVFAFGGAVENPEYLTP 1199
Qy 1194 QGGAAPQHPHPPAPSPAFDNLVYWDQPPPERGAPPSTFKGTPTAENPEYLGDLVVPV 1249
Db 1200 QGGAAPQHPHPPAPSPAFDNLVYWDQPPPERGAPPSTFKGTPTAENPEYLGDLVVPV 1255

RESULT 5

AAW01111 standard; Protein; 1255 AA.

XX AC AAW01111;
DT 01-JAN-1997 (first entry)
XX DE HER-2/neu protein.
XX KW HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
KW breast cancer; ovary cancer; colon cancer; lung cancer;
KW prostate cancer; immunisation; tumour; vaccine; vector.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT Domain 676..1255
FT /label= Intracellular domain
FT /note= "Claimed domain, useful for immunisation"

XX W09630514-A1.

XX 03-OCT-1996.

XX 28-MAR-1996; 96WO-US01689.

XX 31-MAR-1995; 95US-0414417.

XX (UNIW) UNIV WASHINGTON.

XX Cheever MA, Disis ML;

XX WPI; 1996-455361/45.

XX N-PSDB; AAT40739.

XX DNA encoding HER-2-neu poly-peptide(s) - used for prevention or
XX treatment of malignancies with which the HER-2/neu oncogene is
XX associated

XX Claim 2; Page 56-61; 71pp; English.

XX Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is
CC the product of the HER-2/neu oncogene (see also AAT40739). The
CC protein is over-expressed in various cancers, including breast,
CC ovarian, colon, lung and prostate. The intracellular domain of the
CC protein can be used to immunise an animal against a malignancy with
CC which the oncogene is associated. The polypeptide can be produced
CC in transformed host cells for use in immunisation. Alternatively,
CC animal cells are transfected in vivo or ex vivo with a viral vector
CC that directs expression of the polypeptide.
XX SQ Sequence 1255 AA;
Query Match 97.4%; Score 6602; DB 17; Length 1255;
Best Local Similarity 97.2%; Pred No. 0;
Matches 1221; Conservative 6; Mismatches 21; Indels 8; Gaps 2;
Qy 1 MELAAALCRWGLLLALLPPGAASQVCTGTDMLRLPASPTHLDMLRHLHYQGCVVQGNL 60
Db 1 MELAAALCRWGLLLALLPPGAASQVCTGTDMLRLPASPTHLDMLRHLHYQGCVVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIEQVGVVLAHNOVROVPLORLRIVRGTQLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIEQVGVVLAHNOVROVPLQRLRIVRGTQLFEDNYALAVLDNG 120
Qy 121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIORNPOLCYQDTILWKDIFHKNOLA 180
Db 121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIORNPOLCYQDTILWKDIFHKNOLA 180
Qy 181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVCAAGCARCKGLPTDCCHEQC 240
Db 181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVCAAGCARCKGLPTDCCHEQC 240
Qy 241 AAGCTGPKSDCLACLFHNSGICELHCPALVYNTDITFESMPNPGRYTFGASCVTACP 300
Db 241 AAGCTGPKSDCLACLFHNSGICELHCPALVYNTDITFESMPNPGRYTFGASCVTACP 300
Qy 301 YNYLTDVSGCTLVCPLNHOEVTAEDGTORCEKSKPCARVCYGLGMQVTKANSKFIGIT 360
Db 301 YNYLTDVSGCTLVCPLNHOEVTAEDGTORCEKSKPCARVCYGLGMQVTKANSKFIGIT 360
Qy 361 ELEFAGCKKIFGSLAFIPESFDGDPASNTAPLQEQVFEETLEEITGYLYISAWPDSLP 420
Db 361 IQEFAGCKKIFGSLAFIPESFDGDPASNTAPLQEQVFEETLEEITGYLYISAWPDSLP 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGTSWLGSLRSLRGLSGLALIHNTLHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGTSWLGSLRSLRGLSGLALIHNTLHLCFVHTV 480
Qy 481 PWDQLFRNPHQALLHTANRPEDECVGEGACHOLCARGHCWGPQTQVNCVSQFLRGQEC 540
Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGACHOLCARGHCWGPQTQVNCVSQFLRGQEC 540
Qy 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVARC 600
Db 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVARC 600
Qy 601 PSGVKPDLSPYMPIWKFPDEBAGQCPINCTHSCVDLDDKGCAPABQASPLTISAVVG 660
Db 601 PSGVKPDLSPYMPIWKFPDEBAGQCPINCTHSCVDLDDKGCAPABQASPLTISAVVG 660
Qy 661 ILLVVLGVVFGILIKRROQKIRKYMRLRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
Db 661 ILLVVLGVVFGILIKRROQKIRKYMRLRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
Qy 721 RKVKVLGSGAFGQYIKA-----NSKFIGITELVLENTSPKANKEILD EAYVMAGVGS 773
Db 721 RKVKVLGSGAFGQYIKGIMIPDGENVK-IPVAIKVLENTSPKANKEILD EAYVMAGVGS 779
Qy 774 PYVSRLLGICLTSTVQLVTPMPYGCILLDVRNKRGLGSDLLNWCWQIAKMSYLEDV 833
Db 780 PYVSRLLGICLTSTVQLVTPMPYGCILLDVRNKRGLGSDLLNWCWQIAKMSYLEDV 839

Qy	834	RLVHRDLAARNVLKSPHNHVKITDFGLARLLDIDETETHADGGKVP	IKWMALESILRRRF	893
Db	840	RLVHRDLAARNVLKSPHNHVKITDFGLARLLDIDETETHADGGKVP	IKWMALESILRRRF	899
Qy	894	THOSDVMISGYVTWELMTFGAKPYDGI	PAREIPOLLEKGBRLPOPPITCTT	1DVYIMVKW 953
Db	900	THOSDVMISGYVTWELMTFGAKPYDGI	PAREIPOLLEKGBRLPOPPITCTT	1DVYIMVKW 959
Qy	954	MIDSECRPRFRELVSFSERMARDQRFVW	IQNEDLGPASPLDSTFFYRSILLEDDMGDLVD	1013
Db	960	MIDSECRPRFRELVSFSERMARDQRFVW	IQNEDLGPASPLDSTFFYRSILLEDDMGDLVD	1019
Qy	1014	AEELYLVPOQQGFCDDPAPAGAGMVHHRHRSSTRSGGDLJTLGL	EPSEEBAPSPAPSE	1073
Db	1020	AEELYLVPOQQGFCDDPAPAGAGMVHHRHRSSTRSGGDLJTLGL	EPSEEBAPSPAPSE	1079
Qy	1074	GAGSDVFDGDLGMAAGKLOSLPHDPSPLQRYSEDP	TVLPSETDGYVAPLTCSPQPEY	1133
Db	1080	GAGSDVFDGDLGMAAGKLOSLPHDPSPLQRYSEDP	TVLPSETDGYVAPLTCSPQPEY	1139
Qy	1134	VNOPDVRPQPPSPREGPLPAAR	PAGATLERAKTILSPGKNGVVKDVFAGGAVENPEYLTP	1193
Db	1140	VNOPDVRPQPPSPREGPLPAAR	PAGATLERPKTILSPGKNGVVKDVFAGGAVENPEYLTP	1199
Qy	1194	QGGAAPOHPPPPAPAFSPADNLYWQDPP	BERGAPPSTFKGTPTAENPEYI	GLDVPV 1249
Db	1200	QGGAAPOHPPPPAPAFSPADNLYWQDPP	BERGAPPSTFKGTPTAENPEYI	GLDVPV 1255

RESULT 6
AAW92406
ID AAW92406 standard; Protein: 1255 AA.

XX PN US5869445-A.

CC	of this protein is used in a method for eliciting or enhancing an immune response to HER-2/neu protein. The polypeptide can stimulate T cells and B cells to produce an immune response to the HER-2/neu protein. The method can be used for immunisation against a malignancy in which the HER-2/neu oncogene is associated and in the treatment of an existing tumour, or to prevent tumour occurrence or reoccurrence.	XX
CC	Sequence 1255 AA;	
CC	Query Match 97.4%; Score 6602; DB 20; Length 1255;	
CC	Best Local Similarity 97.2%; Pred. No. 0;	
CC	Matches 1221; Conservative 6; Mismatches 21; Indels 8; Gaps 2;	
Qy	1 MELAALCRWGILLALLPPGAASTQVCTGTDWKLRLPASPETHLDMLRHLYQGQCVWGQNL 60	
Db	1 MELAALCRWGILLALLPPGAASTQVCTGTDWKLRLPASPETHLDMLRHLYQGQCVWGQNL 60	
Qy	61 ELTYLPTNASSLFLODIQEVGVYLIAHNQVROVPLQRLIRVRGTQLFDENYALAVLDNG 120	
Db	61 ELTYLPTNASSLFLODIQEVGVYLIAHNQVROVPLQRLIRVRGTQLFDENYALAVLDNG 120	
Qy	121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNOLA 180	
Db	121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNOLA 180	
Qy	181 LTLIDNRSRACHPCSPMKGSRGWSBSSDDCSLRTKTVCGGRCARCKGPLPTDCCHEQC 240	
Db	181 LTLIDNRSRACHPCSPMKGSRGWSBSSDDCSLRTKTVCGGRCARCKGPLPTDCCHEQC 240	
Qy	241 AAGCTGPKHSDCLACLFHNSHSGICELHCPALVTYNTDTFESMPNPEGRTYFGASCVTACP 300	
Db	241 AAGCTGPKHSDCLACLFHNSHSGICELHCPALVTYNTDTFESMPNPEGRTYFGASCVTACP 300	
Qy	301 YNYLSTDVGCTLVCPLNHOEVTAEADGTQRCCKSPCARVCYGLGMQYIKANSKFITG 360	
Db	301 YNYLSTDVGCTLVCPLNHOEVTAEADGTQRCCKSPCARVCYGLGMQYIKANSKFITG 360	
Qy	361 ELSPAGCKKI?GSLAFLPESFDGDPASNTAPLOPEQLQVPEETLEETGYIYISAWPDSLP 420	
Db	361 IQSPAGCKKI?GSLAFLPESFDGDPASNTAPLOPEQLQVPEETLEETGYIYISAWPDSLP 420	
Qy	421 DLSVFQNLQVIRGRIILHNGAYSITLQGLGISWLGRLSRLRGSLALIHNTHLCFVHTV 480	
Db	421 DLSVFQNLQVIRGRIILHNGAYSITLQGLGISWLGRLSRLRGSLALIHNTHLCFVHTV 480	
Qy	481 PDQLFRNPQALHTANRPEDECVBEGSLACHQLCARGHCWGPGPTQCVNCSQFLRQEC 540	
Db	481 PDQLFRNPQALHTANRPEDECVBEGSLACHQLCARGHCWGPGPTQCVNCSQFLRQEC 540	
Qy	541 VECRVULQGLPREYVNAHCLCPHCECPQNGSVTCFGEADQCVACAHYKDPFCVVARC 600	
Db	541 VECRVULQGLPREYVNAHCLCPHCECPQNGSVTCFGEADQCVACAHYKDPFCVVARC 600	
Qy	601 PSGVKPDLSPYMPKPPDEBGAQCPQPCINCTHSCVDLDDKGCPAEQRASPLTSTVSAVVG 660	
Db	601 PSGVKPDLSPYMPKPPDEBGAQCPQPCINCTHSCVDLDDKGCPAEQRASPLTSTVSAVVG 660	
Qy	661 ILLVVVLGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720	
Db	661 ILLVVVLGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720	
Qy	721 RKVKVLGSGAFGOYIKA-----NSKFIGITELVIRENTSPKANKEILDEAYVMAGVGS 773	
Db	721 RKVKVLGSGAFGTYVYIGIWTIPDGENVK-IPVAIKVIRENTSPKANKEILDEAYVMAGVGS 779	
Qy	774 PYVSRLIGICLTSTVOLVTQMPYGCILLDHVRENRGRGLGSODLLNWCQIAKMSYLEDV 833	
Db	780 PYVSRLIGICLTSTVOLVTQMPYGCILLDHVRENRGRGLGSODLLNWCQIAKMSYLEDV 839	
Qy	834 RLVRHDLAARNVLVKSNNHVKITDFGLARLLDIDETEHADGKVPKKNWALSLSLRRRF 893	
Db	840 RLVRHDLAARNVLVKSNNHVKITDFGLARLLDIDETEHADGKVPKKNWALSLSLRRRF 899	

Qy 894 THQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPICTIDVYMWKWCW 953
Db 900 THQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPICTIDVYMWKWCW 959
Qy 954 MIDSECRPRFRELVSERSMARDQRFVVIQNEIDLGPASPLDSTFYRSLLEDMDGLVD 1013
Db 960 MIDSECRPRFRELVSERSMARDQRFVVIQNEIDLGPASPLDSTFYRSLLEDMDGLVD 1019
Qy 1014 ABEYLVPQGFCDPAPAGAGWVHRRHSSTSGGDLTLGLEPSEEEAPRSLAPSE 1073
Db 1020 ABEYLVPQGFCDPAPAGAGWVHRRHSSTSGGDLTLGLEPSEEEAPRSLAPSE 1079
Qy 1074 GAGSDVFDGDLGMAAGLQSLTHDPSPLQRYSEDTVPPLPSETDGYVAPLTCSPOPEY 1133
Db 1080 GAGSDVFDGDLGMAAGLQSLTHDPSPLQRYSEDTVPPLPSETDGYVAPLTCSPOPEY 1139
Qy 1134 VNQPDVPRQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKOVFAFGGAVENPEYLT 1193
Db 1140 VNQPDVPRQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKOVFAFGGAVENPEYLT 1199
Qy 1194 QGGAAPQHPHPPAFSPAFDNLVYNDQPPPERGAPPSTFKGPTAENPEYLGLDV 1249
Db 1200 QGGAAPQHPHPPAFSPAFDNLVYNDQPPPERGAPPSTFKGPTAENPEYLGLDV 1255

RESULT 7
AAB21198
ID AAB21198 standard; protein; 1255 AA.
AC AAB21198;
XX
DT 12-JAN-2001 (first entry)
XX
DE Human HER-2/neu protein.
XX
KW Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;
KW breast cancer; prostate cancer; ovarian cancer; lung cancer;
KW colon cancer.
XX
OS Homo sapiens.
XX
FN WO200044899-A1.
XX
PD 03-AUG-2000.
XX
PF 28-JAN-2000; 2000WO-US02164.
XX
PR 29-JAN-1999; 99US-0117976.
XX
PA (CORI-) CORIXA CORP.
PA (SMIK) SMITHKLINE BEECHAM.
XX
PI Cheever MA, Gheysen D;
XX
DR WPI; 2000-505976/45.
DR N-PSDB; AAA89736.
XX
PT HER-2/neu extracellular domain/phosphorylation domain fusion proteins
PT useful for vaccinating against breast, ovarian, colon, lung and
PT prostate cancers -
XX
PS Claim 52; Fig 7; 128pp; English.
XX
CC The present sequence is the human HER-2/neu protein. It is a member of
CC the tyrosine kinase family of receptor-like glycoproteins and shows
CC homology to the epidermal growth factor receptor (EGFR). It probably
CC plays a part in cell growth and/or differentiation. The HER-2/neu
CC gene is an oncogene. An HER-2/neu fusion protein comprising a
CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
CC domain may be used to treat or prevent cancer by eliciting or
CC enhancing an immune response to the HER-2/neu protein. It may be used
CC to treat malignancies such as breast, ovarian, colon, lung and
CC prostate cancers, and may be used as an antigen to vaccinate against

CC these neoplasias.
XX
SQ Sequence 1255 AA;
Query Match 97.4%; Score 6602; DB 21; Length 1255;
Best Local Similarity 97.2%; Pred. No. 0;
Matches 1221; Conservative 6; Mismatches 21; Indels 8; Gaps 2;
Qy 1 MELAALCRWGLLALLPPGAASQVCTGTDMKRLRLPASPTHLDMLRHLRYGQCOVVOG 60
Db 1 MELAALCRWGLLALLPPGAASQVCTGTDMKRLRLPASPTHLDMLRHLRYGQCOVVOG 60
Qy 61 ELTYLPTNASLSFLQDIQEVQGVYLIHAHQVQVPLQRLRIVRGTLQFEDNYALAVLD 120
Db 61 ELTYLPTNASLSFLQDIQEVQGVYLIHAHQVQVPLQRLRIVRGTLQFEDNYALAVLD 120
Qy 121 DPLNNTTPTVGTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTTLWKDILFHN 180
Db 121 DPLNNTTPTVGTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTTLWKDILFHN 180
Qy 181 LTLIDTNRSRACHPCSPCMCKSGSCWGESSEDCSLTRTVCCAGCARCKGPLPTDCCHE 240
Db 181 LTLIDTNRSRACHPCSPCMCKSGSCWGESSEDCSLTRTVCCAGCARCKGPLPTDCCHE 240
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 300
Qy 301 YNYLSTDVSGCTLVCPLNHOEVTAEDGTQRCCKSPCARVCYGLGHOYIKANSKFTGIT 360
Db 301 YNYLSTDVSGCTLVCPLNHOEVTAEDGTQRCCKSPCARVCYGLGHOYIKANSKFTGIT 360
Qy 361 ELEFAGCKKIFGSLAFIPESFDGDPASNTAPLOEQVETLEEITGYLYISAWPDSLP 420
Db 361 IQBFAGCKKIFGSLAFIPESFDGDPASNTAPLOEQVETLEEITGYLYISAWPDSLP 420
Qy 421 DLSVFQNLQVIRGILHNGAYSILTQGLGISWLSGLSLRELGLALIHNTLHLCFVHTV 480
Db 421 DLSVFQNLQVIRGILHNGAYSILTQGLGISWLSGLSLRELGLALIHNTLHLCFVHTV 480
Qy 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQI CARGHCWGPGPTQVCNCSQPLRQEC 540
Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQI CARGHCWGPGPTQVCNCSQPLRQEC 540
Qy 541 VESCRVLQGLPREVYNARHCLPCHPECPONGSVTCFGEADOCVACAHYKDPFPCVARC 600
Db 541 VESCRVLQGLPREVYNARHCLPCHPECPONGSVTCFGEADOCVACAHYKDPFPCVARC 600
Qy 601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPCAEQASPLTSISAVVG 660
Db 601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPCAEQASPLTSISAVVG 660
Qy 661 ILLVVLGVVFGILIKRROQKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
Db 661 ILLVVLGVVFGILIKRROQKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
Qy 721 RKVKVLGSGAFGOYIKA-----NSKFIGITELVLENTSPKANKILDEAYVMAGVGS 773
Db 721 RKVKVLGSGAFGOYIKGIWIPDGENVK-IPVALKVLRENTSPKANKILDEAYVMAGVGS 779
Qy 774 PYVSRILIGICLTSTVQLVTQMPYGLLDHVRNRRGLSGDQLLNWCQIAGKMSYLEDV 833
Db 780 PYVSRILIGICLTSTVQLVTQMPYGLLDHVRNRRGLSGDQLLNWCQIAGKMSYLEDV 839
Qy 834 RLVRDLAARNVLKSPNHHVKTDFGLARLLDIDETEHADGGKVPKKNWALESLRRRF 893
Db 840 RLVRDLAARNVLKSPNHHVKTDFGLARLLDIDETEHADGGKVPKKNWALESLRRRF 899
Qy 894 THQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPICTIDVYMWKWCW 953
Db 900 THQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPICTIDVYMWKWCW 959
Qy 954 MIDSECRPRFRELVSERSMARDQRFVVIQNEIDLGPASPLDSTFYRSLLEDMDGLVD 1013

|||||
Db 960 MIDSECRPRFELVSEFSRWARDPQRFVVIQNEIDLPASPLDSTFYRSLLEDDMGDLVD 1019
Qy 1014 AEEYLVPQGFCDPAPAGGMVHRRHRSSTRSGGDLTLGLEPSEEEAPRSLAPSE 1073
Db 1020 AEEYLVPQGFCDPAPAGGMVHRRHRSSTRSGGDLTLGLEPSEEEAPRSLAPSE 1079
Qy 1074 GAGSDVFDGLGMAAGLQSLPHDPSPLQRYSEDPTVPLPSETDGVAPLTCSPQPEY 1133
Db 1080 GAGSDVFDGLGMAAGLQSLPHDPSPLQRYSEDPTVPLPSETDGVAPLTCSPQPEY 1139
Qy 1134 VNQDVRPQPPSPREGPLPAARPAATLERAKTLSPGKNGVVDVFAFGGAVENPEYLT 1193
Db 1140 VNQDVRPQPPSPREGPLPAARPAATLERAKTLSPGKNGVVDVFAFGGAVENPEYLT 1199
Qy 1194 QGGAAPQHPHPPAPSPAFDNLVYWDQPPPERGAPPSTFKGPTTAENPEYLGDLVVP 1249
Db 1200 QGGAAPQHPHPPAPSPAFDNLVYWDQPPPERGAPPSTFKGPTTAENPEYLGDLVVP 1255

RESULT 8
AAY84780
ID AAY84780 standard; Protein; 1255 AA.
XX
AC AAY84780;
XX
DT 08-AUG-2000 (first entry)
XX
DE Amino acid sequence of the SPLICE erbB-2 receptor protein.
XX
KW SPLICE erbB-2 receptor protein; cell transformation disorder; cancer;
KW tumor cell proliferation; tissue degeneration; arthropathy;
KW bone resorption; inflammatory disease; degenerative disorder;
KW wound healing.
OS Homo sapiens.
XX
PN W0200020579-A1.
XX
PD 13-APR-2000.
XX
PF 01-OCT-1999; 99WO-CA00912.
XX
PR 02-OCT-1998; 98US-0165192.
XX
PA (UYMC-) UNIV MCMASTER.
XX
PI Muller WJ, Siegel PM;
XX
DR WPI; 2000-303768/26.
DR N-PSDB; AAA14812.
XX
PT Nucleic acid encoding an erbB 2 receptor protein designated SPLICE
PT erbB-2, inhibitors of the protein are useful for treatment of cancer -
XX
PS Claim 3; Fig 2; 60pp; English.
XX
CC The present sequence represents a SPLICE erbB-2 receptor protein. The
CC protein has an in-frame deletion of 16 amino acids, 2 of which are
CC conserved cysteine residues, compared to the unspliced protein. The
CC erbB-2 polynucleotide is used to construct probes for detecting
CC disorders of cell transformation such as cancer. Antibodies to the
CC protein may be used to detect SPLICE erbB-2 in a sample. Agents
CC (e.g. antisense oligonucleotides) which inhibit the expression of
CC SPLICE erbB-2 are useful for reducing tumor cell proliferation and
CC treating cancer. Substances which stimulate SPLICE erbB-2 are useful
CC for treating conditions involving damaged cells including conditions
CC in which degeneration of tissue occurs, such as arthropathy, bone
CC resorption, inflammatory diseases, degenerative disorders of the
CC central nervous system and wound healing.

Sequence 1255 AA;

Query Match 97.4%; Score 6602; DB 21; Length 1255;
Best Local Similarity 97.2%; Pred. No. 0;
Matches 1221; Conservative 6; Mismatches 21; Indels 8; Gaps 2;
Qy 1 MEAAALCRWGLLALLPPGAASQVCTGTDKMLRLPASPTHLDMLRHLVGGCOVVOGNL 60
Db 1 MEAAALCRWGLLALLPPGAASQVCTGTDKMLRLPASPTHLDMLRHLVGGCOVVOGNL 60
Qy 61 ELTYLPTNASLSLQDIQEVQGVYLAHNOVROVPLQRLRIVRGTQLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSLQDIQEVQGVYLAHNOVROVPLQRLRIVRGTQLFEDNYALAVLDNG 120
Qy 121 DPLNNTPTVPGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTTLWKDIFHNQOLA 180
Db 121 DPLNNTPTVPGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTTLWKDIFHNQOLA 180
Qy 181 LTLDITNRSACHPCSPMKGSCWGESSEDCOSLRTVTCAGGCARCKGLPTDCCHEQC 240
Db 181 LTLDITNRSACHPCSPMKGSCWGESSEDCOSLRTVTCAGGCARCKGLPTDCCHEQC 240
Qy 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPPLHNOEVTAEQTCRCEKSKPCARVCYGLGMQYIKANSKFIT 360
Db 301 YNYLSTDVGSCTLVCPPLHNOEVTAEQTCRCEKSKPCARVCYGLGMQYIKANSKFIT 360
Qy 361 ELSEACKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFETLEEITGYLYISAWPSLP 420
Db 361 IQSFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFETLEEITGYLYISAWPSLP 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSLLTLOGLGISWLSGLSLRELGLALIHNTLHCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSLLTLOGLGISWLSGLSLRELGLALIHNTLHCFVHTV 480
Qy 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQLCARHCHWGPGPTQCVNCSQFLRQEC 540
Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQLCARHCHWGPGPTQCVNCSQFLRQEC 540
Qy 541 VEESRVLOGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFCVARC 600
Db 541 VEESRVLOGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFCVARC 600
Qy 601 PSGVKPDLSPYWKPFDEBACQPCINCHSCVDLDDKGCAPAEORASPLTISVAVVG 660
Db 601 PSGVKPDLSPYWKPFDEBACQPCINCHSCVDLDDKGCAPAEORASPLTISVAVVG 660
Qy 661 ILLVWVGVVFGILIKRQOKIRKYTMRLLOQETELVEPLTPSGAMPNQAMRILKETEL 720
Db 661 ILLVWVGVVFGILIKRQOKIRKYTMRLLOQETELVEPLTPSGAMPNQAMRILKETEL 720
Qy 721 RKVKVLGSAFGQYIKA-----NSKFGITELVLRNTSPKANKEILDEAYVMAGVGS 773
Db 721 RKVKVLGSAFGQYIKGIWIPDGENVK-IPVAIKVLRNTSPKANKEILDEAYVMAGVGS 779
Qy 774 PYVSRLLGICLTSTVOLVTOLMPYGLDHDVNRNRLGSGQDLNLMCMQAKGMSYLEDV 833
Db 780 PYVSRLLGICLTSTVOLVTOLMPYGLDHDVNRNRLGSGQDLNLMCMQAKGMSYLEDV 839
Qy 834 RLVRDLAARNVLKSPNVHKITDFGLARLLDIDETEHADGKGVPIKWMALRESILRRRF 893
Db 840 RLVRDLAARNVLKSPNVHKITDFGLARLLDIDETEHADGKGVPIKWMALRESILRRRF 899
Qy 894 THQSDVMSYGVTVWELMTGAKPYDGI PAREIPDLLEKGERLPPQPICTIDVTVMVVKW 953
Db 900 THQSDVMSYGVTVWELMTGAKPYDGI PAREIPDLLEKGERLPPQPICTIDVTVMVVKW 959
Qy 954 MIDSECRPRFELVSEFSRWARDPQRFVVIQNEIDLPASPLDSTFYRSLLEDDMGDLVD 1013
Db 960 MIDSECRPRFELVSEFSRWARDPQRFVVIQNEIDLPASPLDSTFYRSLLEDDMGDLVD 1019
Qy 1014 AEEYLVPQGFCDPAPAGGMVHRRHRSSTRSGGDLTLGLEPSEEEAPRSLAPSE 1073

Db 1140 VNQPDVRRQPPSPREGPLPAARPAAGATLERPKTSLSPKNGVVKDVFAFGGAVENPEYLTP 1199

Qy 1194 QGGAAPQHPHPPAFSPAFDNLYYWDQPPPERGAPPSTFKGTPTAENPEYLGLDVVP 1249

Db 1200 QGGAAPQHPHPPAFSPAFDNLYYWDQPPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 10

ID AAG88267 standard; Protein; 1255 AA.

AC AAG88267;

XX 11-SEP-2001 (first entry)

XX HER2/neu amino acid sequence.

XX Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;

XX immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;

XX tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.

XX Homo sapiens.

XX WO200141787-A1.

XX 14-JUN-2001.

XX 11-DEC-2000; 2000WO-US33591.

XX 10-DEC-1999; 99US-0458299.

XX (EPIM-) EPIMUNE INC.

XX Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;

XX Keogh E;

XX WPI; 2001-374995/39.

XX An isolated prepared HER2/neu epitope useful in a vaccine for inducing

XX cellular immune responses for the prevention and treatment of cancer -

XX Disclosure; Page 15; 1999p; English.

XX The present invention describes isolated prepared HER2/neu epitopes (I).

XX Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is

XX culture in vitro and binds to a complex of an epitope (I), bound to a

XX human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (1)

XX and a second epitope and the peptide is less than 50 contiguous amino

XX acids that have 100% identity with a native peptide sequence of HER2/neu;

XX (3) a vaccine composition (III) comprising (II) and a pharmaceutical

XX excipient; (4) an isolated nucleic acid encoding a peptide comprising

XX (I); and (5) an isolated nucleic acid encoding (II). (I) has cytostatic

XX and immunostimulant activities, and can be used in vaccines. (I), (II)

XX and (III) are useful for inducing cellular immune responses for the

XX prevention and treatment of cancer. (I) and (II) are useful for

XX monitoring or evaluating an immune response to a tumour-associated

XX antigen when incubated with a T lymphocyte sample form a patient and

XX detecting the presence of bound T lymphocyte to (I) or (II). Epitope

XX based vaccines mean that immunosuppressive epitopes that may be present

XX in whole antigens may be avoided. Selected epitopes may be combined to

XX enhance immunogenicity. The possible pathological side effects caused by

XX infectious agents or whole protein antigen is eliminated. The vaccine

XX provides the ability to direct and focus an immune response to multiple

XX selected antigens from the same pathogen. Epitope-based anti-tumour

XX vaccines provides the opportunity to combine epitopes derived from

XX multiple tumour-associated molecules addressing the problem of tumour

XX tumour variability and reducing the likelihood of tumour escape due to

XX antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in

XX the exemplification of the present invention.

XX Sequence 1255 AA;

Query Match

97.4%; Score 6602; DB 22; Length 1255;

Best Local Similarity 97.2%; Pred. No. 0;

Matches 1221; Conservative 6; Mismatches 21; Indels 8; Gaps 2;

Qy 1 MELAAALCRWGLLALLPPGAASQVCTGTDKMLRLPASBETHLDMLRHLYQGCVVQGNL 60

Db 1 MELAAALCRWGLLALLPPGAASQVCTGTDKMLRLPASBETHLDMLRHLYQGCVVQGNL 60

Qy 61 ELTYLPTNASLSFLQDIQEVQGVYLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Db 61 ELTYLPTNASLSFLQDIQEVQGVYLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIORNPOLCYQDTILWKDIHFNKNOLA 180

Db 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIORNPOLCYQDTILWKDIHFNKNOLA 180

Qy 181 LTLIDTNRSRACHPCSPMKSGSRCSSESSEDCCSLTRTVCCAGGCARCKGLPDTDCHEQC 240

Db 181 LTLIDTNRSRACHPCSPMKSGSRCSSESSEDCCSLTRTVCCAGGCARCKGLPDTDCHEQC 240

Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy 301 YNYLSTDVSGCTLVCPHNOEVTAEADGTORCEKCSKPCARVCYGLGMEHLREVRVTSAN 360

Db 301 YNYLSTDVSGCTLVCPHNOEVTAEADGTORCEKCSKPCARVCYGLGMEHLREVRVTSAN 360

Qy 361 ELBFAGCKITFGSLAFLEPESFDGDPASNTAPLOPELOVFPETLEEITGYLIYISAWPDSL 420

Db 361 IQEFAGCKITFGSLAFLEPESFDGDPASNTAPLOPELOVFPETLEEITGYLIYISAWPDSL 420

Qy 421 DLSVFQNLQVIRGIRLHNGAYSILTLQGLGISWGLSLRLSRLSGLALIHNTLHLCFVHTV 480

Db 421 DLSVFQNLQVIRGIRLHNGAYSILTLQGLGISWGLSLRLSRLSGLALIHNTLHLCFVHTV 480

Qy 481 PWDQLFRNPHQALLHTANRPEDECVGEGACHQLCARGHCWGPGTQCVNCSQFLRGOEC 540

Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGACHQLCARGHCWGPGTQCVNCSQFLRGOEC 540

Qy 541 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFCVCAR 600

Db 541 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFCVCAR 600

Qy 601 PSQVKPDLSPMTKPPDEEGACQPCPINCTHSCVDLDDKGCQPAEQASPLTSIVSAVVG 660

Db 601 PSQVKPDLSPMTKPPDEEGACQPCPINCTHSCVDLDDKGCQPAEQASPLTSIVSAVVG 660

Qy 661 ILLVVVLGVVVGILIKRROOKIRKYTMRLLOETELVEPLTPSGAMPNQAMRILKETEL 720

Db 661 ILLVVVLGVVVGILIKRROOKIRKYTMRLLOETELVEPLTPSGAMPNQAMRILKETEL 720

Qy 721 RKVKVLGSGAFQGYIKA-----NSKFGITELVLENTSPKANKEILDEAVVMAGVGS 773

Db 721 RKVKVLGSGAFQGYIKGIWIPGENYK-IPVAIKVLENTSPKANKEILDEAVVMAGVGS 779

Qy 774 PYVSRLLGLCLSTVOLVTLQMPYGLLDHVRNRRGLSQDLNNCMQIAKMSYLEDV 833

Db 780 PYVSRLLGLCLSTVOLVTLQMPYGLLDHVRNRRGLSQDLNNCMQIAKMSYLEDV 839

Qy 834 RLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRRF 893

Db 840 RLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRRF 899

Qy 894 THQSDVMSVGVTVWELMTGAKPYDGI PAREIPDLLEKGERLPPOPICTIDVYMWVKW 953

Db 900 THQSDVMSVGVTVWELMTGAKPYDGI PAREIPDLLEKGERLPPOPICTIDVYMWVKW 959

Qy 954 MIDSECRPRELVSFSSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLEDDDDMDGLVD 1013

Db 960 MIDSECRPRELVSFSSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLEDDDDMDGLVD 1019

Qy 1014 ABEYLVPQGGFTCPDPAPGAGGMVHRHRSSTSTRSGGDLTLGLEPSEEPASPLAPSE 1073

Db 1020 ABEYLVPQGGFFCPDPAPAGAGWVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSE 1079
Qy 1074 GAGSDVFDGLGMAAGLQSLPTHDSPLORYSEDPTVPLPSETDGVVAPLTCSPQPEY 1133
Db 1080 GAGSDVFDGLGMAAGLQSLPTHDSPLORYSEDPTVPLPSETDGVVAPLTCSPQPEY 1139
Qy 1134 VNQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVYVVDVAFGGAVENPEYLTP 1193
Db 1140 VNQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVYVVDVAFGGAVENPEYLTP 1199
Qy 1194 QGGAAPQPPPPAPSPAFDNLYYWDQDPPERGAPPSTFKGPTTAENPEYGLDVPV 1249
Db 1200 QGGAAPQPPPPAPSPAFDNLYYWDQDPPERGAPPSTFKGPTTAENPEYGLDVPV 1255

RESULT 11

AAE24067
ID AAE24067 standard; Protein; 1255 AA.

XX AC AAE24067;

XX DT 23-SEP-2002 (first entry)

XX DE Human Her-2 protein.

XX KW Human; Her-2; epidermal growth factor receptor 2; infection; cancer;
XX KW hyperproliferative disorder; prophylaxis; inflammation; antisense;
XX KW tumour; gene therapy; phosphorothioate backbone.

XX OS Homo sapiens.

XX PN WO200222636-A1.

XX PD 21-MAR-2002.

XX PF 12-SEP-2001; 2001WO-US28572.

XX PR 15-SEP-2000; 2000US-0663834.

XX PA (ISIS-) ISIS PHARM INC.

XX PI Bennett CF, Cowser LM;

XX DR WPI; 2002-471192/50.

XX DR N-PSDB; RAD38904.

XX PT Novel antisense oligonucleotide which modulates the expression of Human
XX PT Epidermal Growth Factor receptor, Her2, is useful for treating tumors
XX PT inflammation or to prevent infection in humans -

XX PS Example 13; Page 95-107; 116pp; English.

XX CC The invention relates to antisense compounds targetted to a nucleic
XX CC acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)
XX CC that specifically hybridises with and inhibits the expression of Her2.
XX CC Antisense compounds of the invention are used for treating diseases or
XX CC conditions associated with Her2 such as hyperproliferative disorders
XX CC e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,
XX CC neural or cardiac cancer. They are also useful prophylactically e.g.
XX CC to prevent or delay infection, inflammation and tumour formation. The
XX CC invention is also used in gene therapy. The present sequence is human
XX CC Her-2 protein.

XX SQ Sequence 1255 AA;

Query Match 97.4%; Score 6602; DB 23; Length 1255;

Best Local Similarity 97.2%; Pred. No. 0;

Matches 1221; Conservative 6; Mismatches 21; Indels 8; Gaps 2;

Qy 1 MELAALCRWGLLALLPPGAASCTGCTDKMLPLPASPETHLDMRLHYGGCQVQGNL 60

Db 1 MELAALCRWGLLALLPPGAASCTGCTDKMLPLPASPETHLDMRLHYGGCQVQGNL 60

Qy 61 ELTYLPTNASLSLQDIQEVQGYVLIHAHQVQVPLQRLAIRVGTQOLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSLQDIQEVQGYVLIHAHQVQVPLQRLAIRVGTQOLFEDNYALAVLDNG 120
Qy 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIORNPOLCYQDTILWKDIFHKNOLA 180
Db 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIORNPOLCYQDTILWKDIFHKNOLA 180
Qy 181 LTLIDNRSRACHPCSPMKGSRGWESSEDOSLRTVTCAGGCARCKGLPDTDCCHQEC 240
Db 181 LTLIDNRSRACHPCSPMKGSRGWESSEDOSLRTVTCAGGCARCKGLPDTDCCHQEC 240
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Qy 301 YNYLSTDVSGCTLVCPHNOEVTAEQGTQRCCKSPCARVCYGLGMOYTKANSKFTGIT 360
Db 301 YNYLSTDVSGCTLVCPHNOEVTAEQGTQRCCKSPCARVCYGLGMOYTKANSKFTGIT 360
Qy 361 ELSPAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVPEETLEEITGYLYISAWPDSL 420
Db 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVPEETLEEITGYLYISAWPDSL 420
Qy 421 DLSVFQNLQVIRGRIHLHNGAYSLTLOGLGISWLSRLSRLSGLALIHNTLHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRIHLHNGAYSLTLOGLGISWLSRLSRLSGLALIHNTLHLCFVHTV 480
Qy 481 PWDQLFRNPHQALLHTANRPEDECVEGLACHQLCARHCHGWGPGTQCVNCSQFLRQEC 540
Db 481 PWDQLFRNPHQALLHTANRPEDECVEGLACHQLCARHCHGWGPGTQCVNCSQFLRQEC 540
Qy 541 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADOCVACAHYKDPFCVARC 600
Db 541 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADOCVACAHYKDPFCVARC 600
Qy 601 PSGVKPDLSPYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCFAEORASPLTSTVSAVVG 660
Db 601 PSGVKPDLSPYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCFAEORASPLTSTVSAVVG 660
Qy 661 ILLVVLGVVFGILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
Db 661 ILLVVLGVVFGILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
Qy 721 RKVKVLGSGAFGOYIKA-----NSKFTGITELVRENTSPKANKEILDEAYVMAGVGS 773
Db 721 RKVKVLGSGAFGOYIKWIPDGENVK-IPVAIKVIRENTSPKANKEILDEAYVMAGVGS 779
Qy 774 PYVSRLIGLICLTSTVQLVTQLMFYGCLLDHVRENRLGSLQDILLNWCMIKAGMSYLEDV 833
Db 780 PYVSRLIGLICLTSTVQLVTQLMFYGCLLDHVRENRLGSLQDILLNWCMIKAGMSYLEDV 839
Qy 834 RLVRDLAARNVLVKSNNHVKITDFGLARLLDIDETEHADGKVPKIMWALSILRRRF 893
Db 840 RLVRDLAARNVLVKSNNHVKITDFGLARLLDIDETEHADGKVPKIMWALSILRRRF 899
Qy 894 THQSDVMSYGVTVWELMTGAKYDGIIPAREIPDLLEKGERLPQPPICITDVMVMVKCW 953
Db 900 THQSDVMSYGVTVWELMTGAKYDGIIPAREIPDLLEKGERLPQPPICITDVMVMVKCW 959
Qy 954 MIDSECRPRELIVSEFSRMARDPQRFVITQNEDLGPASPLDSTFYRSLLEDDMDGLVD 1013
Db 960 MIDSECRPRELIVSEFSRMARDPQRFVITQNEDLGPASPLDSTFYRSLLEDDMDGLVD 1019
Qy 1014 ABEYLVPQGGFFCPDPAPAGAGWVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSE 1073
Db 1020 ABEYLVPQGGFFCPDPAPAGAGWVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSE 1079
Qy 1074 GAGSDVFDGLGMAAGLQSLPTHDSPLORYSEDPTVPLPSETDGVVAPLTCSPQPEY 1133
Db 1080 GAGSDVFDGLGMAAGLQSLPTHDSPLORYSEDPTVPLPSETDGVVAPLTCSPQPEY 1139
Qy 1134 VNQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVYVVDVAFGGAVENPEYLTP 1199

Db 1020 AEEYLVPOQGFCDPAPGAGWVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSE 1079
 Qy 1074 GAGSDVPGDGLGMAAGLQSLPTHDPSPLOQYSEDPTVPLPSTGVDGYVAPLTCSPOPEY 1133
 Db 1080 GAGSDVPGDGLGMAAGLQSLPTHDPSPLOQYSEDPTVPLPSTGVDGYVAPLTCSPOPEY 1139
 Qy 1134 VNQPDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVDVFAFGGAVENPEYLTTP 1193
 Db 1140 VNQPDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVDVFAFGGAVENPEYLTTP 1199
 Qy 1194 QGGAAPQHPPPAPSPAFDNLVYWDQDPPERGAPSPSTFKGPTTAENPEYILGVDVVP 1249
 Db 1200 QGGAAPQHPPPAPSPAFDNLVYWDQDPPERGAPSPSTFKGPTTAENPEYILGVDVVP 1255

RESULT 13

AA51143
 ID AA51143 standard; Protein; 1255 AA.

XX AC AA51143;

XX DT 17-JUN-2002 (first entry)

XX DE Human Her-2/neu oncogene-encoded p185 glycoprotein.

XX KW Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;

XX KW tyrosine kinase; receptor; c-erbB2; gene therapy.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT Domain 1..653

XX FT /note= "extracellular domain"

XX FT Domain 676..1255

XX FT /note= "intracellular domain"

XX FT Domain 990..1255

XX FT /note= "phosphorylation domain"

XX PN WO200212341-A2.

XX PD 14-FEB-2002.

XX XX 03-AUG-2001; 2001WO-US24283.

XX XX 03-AUG-2000; 2000US-0632507.

XX XX (CORI-) CORIXA CORP.

XX XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX XX Cheever MA, Gheysen D;

XX XX WPI; 2002-241743/29.

XX XX N-PSDB; ABA92250.

XX Her-2/neu fusion protein for treating or preventing cancer by eliciting

XX PT or enhancing an immune response to the protein, has Her-2/neu

XX PT extracellular domain fused to Her-2/neu intracellular or

XX PT phosphorylation domain

XX PS Claim 68; Fig 7; 141pp; English.

XX The present sequence is that of human Her-2/neu (p185 glycoprotein
 CC or c-erbB2), an oncogenic self-protein and target for anti-cancer
 CC vaccines. The Her-2/neu gene is amplified and p185 is overexpressed
 CC in a variety of cancers, including breast, ovarian, colon, lung and
 CC prostate cancer. Her-2/neu is a member of the tyrosine kinase
 CC family of receptor-like glycoproteins. It comprises an extracellular
 CC domain with homology to the epidermal growth factor receptor
 CC (EGFR), a highly hydrophobic transmembrane domain and a C-terminal
 CC intracellular domain that also shows homology to EGFR. Its
 CC overexpression correlates with a poor prognosis in breast and
 CC ovarian cancers. The invention provides Her-2/neu fusion
 CC proteins, nucleic acids encoding them, viral vectors, and vaccines

CC comprising the fusion proteins or nucleic acid molecules. In
 CC preferred fusion proteins, the extracellular domain of a Her-2/neu
 CC protein is fused to a Her-2/neu intracellular domain or
 CC phosphorylation domain (or its DeltaPD fragment). An immune
 CC response to Her-2/neu protein is elicited or enhanced by
 CC administering the fusion protein in the form of a vaccine, or by
 CC transfecting cells of an animal *ex vivo* with a nucleic acid
 CC encoding the fusion protein, and delivering the transfected cells
 CC to the animal. The fusion proteins, nucleic acids, and isolated
 CC specific T-cells are useful for inhibiting the development of a
 CC cancer, especially breast, ovarian, colon, lung or prostate cancer
 CC in a patient. T cells that specifically react with a Her-2/neu
 CC fusion protein can be used to remove tumour cells from a sample in
 CC order to inhibit the development of cancer in a patient.

XX SQ Sequence 1255 AA;

Query Match 97.4%; Score 6602; DB 23; Length 1255;

Best Local Similarity 97.2%; Pred. No. 0;

Matches 1221; Conservative 6; Mismatches 21; Indels 8; Gaps 2;

Qy 1 MELAALCRWGLLALLPPGAASTQVCTGDMKRLRPASPETHLDMRLHLYQGCVVQGNL 60

Db 1 MELAALCRWGLLALLPPGAASTQVCTGDMKRLRPASPETHLDMRLHLYQGCVVQGNL 60

Qy 61 ELTYLPTNASLSFLQDIQEVQGVVLAHNOVQVPLQRLRVGTQTFEDNYALAVLDNG 120

Db 61 ELTYLPTNASLSFLQDIQEVQGVVLAHNOVQVPLQRLRVGTQTFEDNYALAVLDNG 120

Qy 121 DPLNTPPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHNKOLA 180

Db 121 DPLNTPPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHNKOLA 180

Qy 181 LTLIDNRSRACHPCSPMKSGRCWGESSEDCOSLTRTVCAGCARGKGLPTDCCHEQC 240

Db 181 LTLIDNRSRACHPCSPMKSGRCWGESSEDCOSLTRTVCAGCARGKGLPTDCCHEQC 240

Qy 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy 301 YNYLSTDVGSCTLVCPPLHNOEVTAEQGTQRCCKSPCARVCYGLGMQYIKANSKFTGIT 360

Db 301 YNYLSTDVGSCTLVCPPLHNOEVTAEQGTQRCCKSPCARVCYGLGMQYIKANSKFTGIT 360

Qy 361 ELEFAGCKKIFGSLAFPLPESFDGDPASNTAPLOPEQLQVFEETLEETGYLYISAWPDSL 420

Db 361 IQEFAGCKKIFGSLAFPLPESFDGDPASNTAPLOPEQLQVFEETLEETGYLYISAWPDSL 420

Qy 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGISWLGRLSLRELGSGLALIHNTLHLCFVHTV 480

Db 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGISWLGRLSLRELGSGLALIHNTLHLCFVHTV 480

Qy 481 PWQLFRNPQALLHTANRPEDECVEGLACHOLCARGHCHWGPGTQCVNCSQFLRGQEC 540

Db 481 PWQLFRNPQALLHTANRPEDECVEGLACHOLCARGHCHWGPGTQCVNCSQFLRGQEC 540

Qy 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVARC 600

Db 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVARC 600

Qy 601 PSGVKPDLSTYMPWKPPDEBACQPCPINCTHSCVDLDDKGCAPAEQASPLTSTVSAVVG 660

Db 601 PSGVKPDLSTYMPWKPPDEBACQPCPINCTHSCVDLDDKGCAPAEQASPLTSTVSAVVG 660

Qy 661 ILLVVLGVVFGILIKRQKIRKYTMRLRLQETELVEPLTPSGAMPNQAMRILKETEL 720

Db 661 ILLVVLGVVFGILIKRQKIRKYTMRLRLQETELVEPLTPSGAMPNQAMRILKETEL 720

Qy 721 RKVKVLGSGAFGQYIKA-----NSKFTGITELVLRNTSPKANKEILDEAYVMAGVGS 773

Db 721 RKVKVLGSGAFGQYIKWTIPDGENVK-IPVAIKVLRNTSPKANKEILDEAYVMAGVGS 779

QY 774 PYVSRLLGICLTSTVQLTQMPYGCCLLDHVRENRLGSDLLNWCQIAKMSYLEDV 833
Db 780 PYVSRLLGICLTSTVQLTQMPYGCCLLDHVRENRLGSDLLNWCQIAKMSYLEDV 839
QY 834 RLVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMWALESIILRRRF 893
Db 840 RLVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMWALESIILRRRF 899
QY 894 THQSDVMSYGVVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVVMVWKW 953
Db 900 THQSDVMSYGVVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVVMVWKW 959
QY 954 MIDSECPRRRELVSERMRARDQRFVVIQNEIDLGPASPLDSTFYRSLLDEDDMGDLVD 1013
Db 960 MIDSECPRRRELVSERMRARDQRFVVIQNEIDLGPASPLDSTFYRSLLDEDDMGDLVD 1019
QY 1014 AEEVLVPQQGFFCPDPAPGAGMVHRRHSSTRSGGDLTLGLEPSEERAPRSLAPSE 1073
Db 1020 AEEVLVPQQGFFCPDPAPGAGMVHRRHSSTRSGGDLTLGLEPSEERAPRSLAPSE 1079
QY 1074 GAGSDVFDGLGMAAKGLOSLPTHDPSPLOQYSEDPTVPLPSETDGYVAPLTCSPQPEY 1133
Db 1080 GAGSDVFDGLGMAAKGLOSLPTHDPSPLOQYSEDPTVPLPSETDGYVAPLTCSPQPEY 1139
QY 1134 VNQPDVRPOPSPREGPLPAARPAAGATILERAKTLSPGKNGVVKDVFAGGAVENPEYLT 1193
Db 1140 VNQPDVRPOPSPREGPLPAARPAAGATILERAKTLSPGKNGVVKDVFAGGAVENPEYLT 1199
QY 1194 QGGAAPQHPHPPAFSPAFDNLVYWDQPPERPAPSTFKGTPTAENPEYLGLDVVP 1249
Db 1200 QGGAAPQHPHPPAFSPAFDNLVYWDQPPERPAPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 14
AAU77114
ID AAU77114 standard; Protein; 1255 AA.
XX AC AAU77114;
XX DT 05-JUN-2002 (first entry)
XX DE Human Her-2/neu polypeptide.
XX KW Human; Her-2/neu; cytostatic; haematological malignancy; CML;
KW acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;
KW chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;
KW Hodgkin's lymphoma; T cell therapy.
XX OS Homo sapiens.
XX PN WO200213847-A2.
XX PD 21-FEB-2002.
XX PF 13-AUG-2001; 2001WO-US25408.
XX PR 14-AUG-2000; 2000US-0638280.
XX PR 28-SEP-2000; 2000US-0675904.
XX PA (CORI-) CORIXA CORP.
XX PI Gaiger A, Cheever MA, Hand-zimmermann S;
XX WPI; 2002-280741/32.
XX DR N-PSDB; ABK10730.
XX PT Inhibiting haematological malignancy development by administering
PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
PT encoding the polypeptide, or antigen presenting cells expressing the
PT polypeptide.
XX PS Disclosure; Page 71-74; 74pp; English.
XX

CC The invention relates to a method for inhibiting development of
CC haematological malignancy in a patient by administering a polypeptide
CC comprising an immunogenic portion of Her-2/neu or a polynucleotide
CC encoding the polypeptide. Antigen presenting cells that express the
CC protein can also be administered. The sequences are used for inhibiting
CC development of haematological malignancy, such as acute myelogenous
CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
CC lymphoma. This sequence represents the human Her-2/neu polypeptide.
XX

SQ Sequence 1255 AA;

Query Match 97.4%; Score 6602; DB 23; Length 1255;
Best Local Similarity 97.2%; Pred. No. 0;
Matches 1221; Conservative 6; Mismatches 21; Indels 8; Gaps 2;
QY 1 MELAALCRWGLLALLPPGAASQVCTGTDMLKRLPASPTHLDMLRHLYOGQOVVGNL 60
Db 1 MELAALCRWGLLALLPPGAASQVCTGTDMLKRLPASPTHLDMLRHLYOGQOVVGNL 60
QY 61 ELYTLPTNASLSFLQDIQEVQGVLIQHNQVQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
Db 61 ELYTLPTNASLSFLQDIQEVQGVLIQHNQVQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
QY 121 DPLNNTPTVGTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIHFNKNOLA 180
Db 121 DPLNNTPTVGTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIHFNKNOLA 180
QY 181 LTLIDNTRSPACHPCSPMCKSGSCWGESSEDCOSLTRVCAGGCARCKGLPTDCCHEQC 240
Db 181 LTLIDNTRSPACHPCSPMCKSGSCWGESSEDCOSLTRVCAGGCARCKGLPTDCCHEQC 240
QY 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPLNHQVTAEDGTQCEKSKPCARVCYGLGMYIKANSKFTGIT 360
Db 301 YNYLSTDVGSCTLVCPLNHQVTAEDGTQCEKSKPCARVCYGLGMYIKANSKFTGIT 360
QY 361 ELEFAGCKIFGSLAPLPSFDGDPASNTAPLOPEQLQVPELLEEITGYLIYISAWPDSL 420
Db 361 IQEFAGCKIFGSLAPLPSFDGDPASNTAPLOPEQLQVPELLEEITGYLIYISAWPDSL 420
QY 421 DLSVFQNLQVIRGRILHNGAYSITLQGLGTSWLGSLRLSRELGSGLALIHNTLHCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSITLQGLGTSWLGSLRLSRELGSGLALIHNTLHCFVHTV 480
QY 481 PWDOLFNPQHALLHTANRPEDECVGEGGLACHOLCARGHCWGPPTQCVNCSQFLRQEC 540
Db 481 PWDOLFNPQHALLHTANRPEDECVGEGGLACHOLCARGHCWGPPTQCVNCSQFLRQEC 540
QY 541 VEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFVCVARC 600
Db 541 VEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFVCVARC 600
QY 601 PSGVKPDLSPYMPIKPPDEGACQPCPINCTHSCVDLDDKGCPEAORASPLTISVAVWG 660
Db 601 PSGVKPDLSPYMPIKPPDEGACQPCPINCTHSCVDLDDKGCPEAORASPLTISVAVWG 660
QY 661 ILLVVLGVVVFGLIKRRQOKIRKYTMRLQLQETELVEPLTPSGAMPNQAMRILKETEL 720
Db 661 ILLVVLGVVVFGLIKRRQOKIRKYTMRLQLQETELVEPLTPSGAMPNQAMRILKETEL 720
QY 721 RKVVLGSGAFGQYIYA-----NSKFIGITELVIRENTSPKANKEILD EATVMAVGVS 773
Db 721 RKVVLGSGAFGQYIYA-----NSKFIGITELVIRENTSPKANKEILD EATVMAVGVS 779
QY 774 PYVSRLLGICLTSTVQLTQMPYGCCLLDHVRENRLGSDLLNWCQIAKMSYLEDV 833
Db 780 PYVSRLLGICLTSTVQLTQMPYGCCLLDHVRENRLGSDLLNWCQIAKMSYLEDV 839
QY 834 RLVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMWALESIILRRRF 893

||||| 840 RLVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRF 899
QY THOSDWSYGVYVWELMTFGAKPDGIPAREIPDLLEKGERLPPOPICTIDVYIMVKW 953
DB THOSDWSYGVYVWELMTFGAKPDGIPAREIPDLLEKGERLPPOPICTIDVYIMVKW 959
QY MIDSECRPRELVSFSESRMARDPQRFVVIQNEBGLPASPLDSTFYRSLLEDDDMGLVD 1013
DB MIDSECRPRELVSFSESRMARDPQRFVVIQNEBGLPASPLDSTFYRSLLEDDDMGLVD 1019
QY ABEYLVPQOQFFCPDPAPAGGMVHHRSSSTRSGGDLTLGLEPSEERAPRSLAPSE 1073
DB ABEYLVPQOQFFCPDPAPAGGMVHHRSSSTRSGGDLTLGLEPSEERAPRSLAPSE 1079
QY GAGSDVFDGLGMAAGKLOSLPHDPSPLORYSEDTVPPLPSTDGYVAPLTCSPQPEY 1133
DB GAGSDVFDGLGMAAGKLOSLPHDPSPLORYSEDTVPPLPSTDGYVAPLTCSPQPEY 1139
QY VNQPDVPRQPPSPREGPLPAARPAAGATLERAKTSLPGKNGVVKDVFAGGAVENPEYLT 1193
DB VNQPDVPRQPPSPREGPLPAARPAAGATLERAKTSLPGKNGVVKDVFAGGAVENPEYLT 1199
QY QGGAAPQHPHPPAFSPAFDNLYYWDQPPPERGAPSTFKGTPTAENPEYGLDVPV 1249
DB QGGAAPQHPHPPAFSPAFDNLYYWDQPPPERGAPSTFKGTPTAENPEYGLDVPV 1255

RESULT 15

AAR39568
ID AAR39568 standard; Protein; 1433 AA.
XX AC AAR39568;
XX DT 07-FEB-1994 (first entry)
XX DE Sequence of c-erbB-2 tumour antigen.
XX KW Tumour antigen; c-erbB-2; glycoprotein.
XX OS Homo sapiens.
XX PN WO9316185-A.
XX PD 19-AUG-1993.
XX PF 05-FEB-1993; 93WO-US01055.
XX PR 06-FEB-1992; 92US-0831967.
XX PA (CETU) CETUS ONCOLOGY CORP.
XX PA (CREA-) CREATIVE BIOMOLECULES INC.
XX PT Houston LL, Huston JS, Oppermann H, Ring DB;
XX PI WPI; 1993-272889/34.
XX DR N-PSDB; AAQ46083.
XX DR New single chain Fv polypeptide binding to C-erbB-2 tumour
XX PT antigen - for imaging or treating breast or ovarian cancer etc.
XX PS Disclosure; pages 48-54; 87pp; English.
XX CC c-erbB-2 refers to a protein antigen expressed on the surface of
XX CC tumour cells. Such as breast and ovarian tumour cells, which is an
XX CC approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric
XX CC pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39583 represents
XX CC the location of a stop codon in AAQ46083.
XX SQ Sequence 1433 AA;

Query Match 96.8%; Score 6559; DB 14; Length 1433;
Best Local Similarity 96.7%; Pred. No. 0;

Matches 1214; Conservative 8; Mismatches 26; Indels 8; Gaps 2;
QY 1 METAAALCRWGLLALLPPGAASTQVCTGTDKMLRLPASPTHLDMLRHLHLYQGQVVOGNL 60
DB 1 METAAALCRWGLLALLPPGAASTQVCTGTDKMLRLPASPTHLDMLRHLHLYQGQVVOGNL 60
QY 61 ELTYLPTNASLSFLQDIQEVQGVYLAHNOVQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSFLQDIQEVQGVYLAHNOVQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
QY 121 DPLNTPVTGASPGGLRELQRLSLTEILKGGVLIORNPOLCYQDTILWKDI FHNKQLA 180
DB 121 DPLNTPVTGASPGGLRELQRLSLTEILKGGVLIORNPOLCYQDTILWKDI FHNKQLA 180
QY 181 LTLIDNRSRACHPCSPMKSGSESSDCSLTRTVCCAGCARCKGKPLPTDCCHEQC 240
DB 181 LTLIDNRSRACHPCSPMKSGSESSDCSLTRTVCCAGCARCKGKPLPTDCCHEQC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVSGCTLVCPLNHOEVTABDGTQRCCKSKPCARVCYGLGMQYIKANSKFTGIT 360
DB 301 YNYLSTDVSGCTLVCPLNHOEVTABDGTQRCCKSKPCARVCYGLGMQYIKANSKFTGIT 360
QY 361 ELFAFACKTIFGSLAFPLPESFDGDPASNTAPLOEOLQVPELLEETIGYLYISAWPDSLP 420
DB 361 ELFAFACKTIFGSLAFPLPESFDGDPASNTAPLOEOLQVPELLEETIGYLYISAWPDSLP 420
QY 421 DLSVFQNLQVIRGRILHNGAYSILTLOGLGISWLSGLSLRELGSGLALIHNTLHCFVHTV 480
DB 421 DLSVFQNLQVIRGRILHNGAYSILTLOGLGISWLSGLSLRELGSGLALIHNTLHCFVHTV 480
QY 481 PWDQLFRNPHQALLHTANRDECEVGEGLACHOLCARGHCWGPGTQCVNCSQFLRQEC 540
DB 481 PWDQLFRNPHQALLHTANRDECEVGEGLACHOLCARGHCWGPGTQCVNCSQFLRQEC 540
QY 541 VEECRVLQGLPREYVNAHCLPCHPCOPONGSVTCFGEADOCVCAHAKYKDPFCVARC 600
DB 541 VEECRVLQGLPREYVNAHCLPCHPCOPONGSVTCFGEADOCVCAHAKYKDPFCVARC 600
QY 601 PSGVKPDLSPYMPKPPDEEGACQPCPINCTHSCVDLDDKGCPEAORASPLTISVAVVG 660
DB 601 PSGVKPDLSPYMPKPPDEEGACQPCPINCTHSCVDLDDKGCPEAORASPLTISVAVVG 660
QY 661 ILLVVLGVVFGILIKRROKIRKYTMRLLOETELVEPLTPSGAMPNQAMRILKETEL 720
DB 661 ILLVVLGVVFGILIKRROKIRKYTMRLLOETELVEPLTPSGAMPNQAMRILKETEL 720
QY 721 RKVKVLGSGAFQGYIKA-----NSKFIGITELVIRENTSPKANKEILDEAYVMAGVGS 773
DB 721 RKVKVLGSGAFQGYIKA-----NSKFIGITELVIRENTSPKANKEILDEAYVMAGVGS 779
QY 774 PYVSRLGICLTSTVOLVTQMPYGLLDHVRNRRGLSGQDLLNWCQIAKMSYLEDV 833
DB 774 PYVSRLGICLTSTVOLVTQMPYGLLDHVRNRRGLSGQDLLNWCQIAKMSYLEDV 839
QY 834 RLVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRF 893
DB 834 RLVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRF 899
QY 894 THOSDWSYGVYVWELMTFGAKPDGIPAREIPDLLEKGERLPPOPICTIDVYIMVKW 953
DB 894 THOSDWSYGVYVWELMTFGAKPDGIPAREIPDLLEKGERLPPOPICTIDVYIMVKW 959
QY 954 MIDSECRPRELVSFSESRMARDPQRFVVIQNEBGLPASPLDSTFYRSLLEDDDMGLVD 1013
DB 954 MIDSECRPRELVSFSESRMARDPQRFVVIQNEBGLPASPLDSTFYRSLLEDDDMGLVD 1019
QY 1014 ABEYLVPQOQFFCPDPAPAGGMVHHRSSSTRSGGDLTLGLEPSEERAPRSLAPSE 1073
DB 1014 ABEYLVPQOQFFCPDPAPAGGMVHHRSSSTRSGGDLTLGLEPSEERAPRSLAPSE 1079

Qy	1074	GAGSDVFDGDLGMAAGLQSLPTHDSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEY	1133
Db	1080	GAGSDVFDGDLGMAAGLQSLPTHDSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEY	1139
Qy	1134	VNQPDVRFQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAGGAVENPEYLTP	1193
Db	1140	VNQPDVRFQPPSPREGPLPAARPAAGATLERPKTILSPGKNGVVKDVFAGGAVENPEYLTP	1199
Qy	1194	QGGAAPQHPHPPAFSPAFDNLYYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVVP	1249
Db	1200	QGGAAPQHPHPPAFSPAFDNLYYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVVP	1255

Search completed: July 22, 2003, 09:16:55
Job time : 42.7573 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:25:54 ; Search time 22.9062 Seconds
(without alignments)
5267.077 Million cell updates/sec

Title: SEQ4-5-25-14
Perfect score: 6812
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVPV 1255

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 73.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6694	98.3	1255	1 A24571	protein-tyrosine k
2	5879.5	86.3	1254	2 148161	p-185 precursor -
3	5876	86.3	1260	1 TVRTNU	protein-tyrosine k
4	3132	46.0	1210	1 GQHUE	epidermal growth f
5	3111	45.7	1210	2 A53183	epidermal growth f
6	3086.5	45.3	1223	1 TVCHLV	epidermal growth f
7	2968.5	43.6	1308	2 A47253	epidermal growth f
8	2689	39.5	1166	1 S06142	protein-tyrosine k
9	2408	35.3	1342	2 A36223	kinase-related tra
10	2327.5	34.2	1339	2 JC4387	epidermal growth f
11	1786.5	25.9	698	1 TVFVLV	protein-tyrosine k
12	1703	25.0	604	1 TVYUUV	protein-tyrosine k
13	1652.5	24.3	1330	1 GQFFE	epidermal growth f
14	1647	24.2	544	2 S35745	protein-tyrosine k
15	1640	24.1	545	2 S00727	kinase-related tra
16	1623	23.8	540	2 B4776	protein-tyrosine k
17	1621	23.8	540	1 TVFVEB	protein-tyrosine k
18	1505	22.1	644	2 A36325	epidermal growth f
19	1297	19.0	1323	2 E88257	protein let-23 [im
20	1297	19.0	1374	2 S70712	protein-tyrosine k
21	1189	17.5	1369	2 S70713	protein-tyrosine k
22	1175	17.2	1717	1 A45558	epidermal growth f
23	1118	16.4	527	2 A42032	epidermal growth f
24	953.5	14.0	843	2 A27131	epidermal growth f
25	806.5	11.8	346	2 S13807	protein-tyrosine k
26	754.5	11.1	311	2 S13808	protein-tyrosine k
27	724	10.6	1363	2 T43220	insulin-like growt
28	702	10.3	1382	1 INHUR	insulin receptor p
29	699.5	10.3	1607	2 T43212	insulin-like growt

ALIGNMENTS

RESULT 1

A24571
protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human
N;Alternate names: c-erb-B-2 protein precursor; kinase-related transforming protein erb
C;Species: Homo sapiens (man)
C;Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999
C;Accession: A24571; A25491; A44188; B44188; I59509; I57622
R;Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T.;
Nature 319, 230-234, 1986
A;Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth
A;Reference number: A24571; MUID:86118663; PMID:3003577
A;Accession: A24571
A;Molecule type: mRNA
A;Residues: 1-1255 <YAM>
A;Cross-references: GB:X03363; NID:g31197; PIDN:CAA27060.1; PID:g31198
R;Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.
Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985
A;Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epider
A;Reference number: A25491; MUID:86016729; PMID:2995967
A;Accession: A25491
A;Molecule type: DNA
A;Residues: 737-1031 <SHM>
A;Cross-references: GB:M1767; NID:gl82163; PIDN:AAA35808.1; PID:g553282
R;Coussens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seesburg,
Science 230, 1132-1139, 1985
A;Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chromo
A;Reference number: A44188; MUID:86070181; PMID:2999974
A;Accession: A44188
A;Molecule type: DNA
A;Residues: 740-910 <COU1>
A;Cross-references: GB:M12036; PIDN:AAA35978.1; PID:g183989
A;Accession: B44188
A;Molecule type: mRNA
A;Residues: 1-517, 'RALL', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>
A;Cross-references: GB:M11730; NID:gl83986
R;King, C.R.; Kraus, M.E.; Aaronson, S.A.
Science 229, 974-976, 1985
A;Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.
A;Reference number: I59509; MUID:85272597; PMID:2992089
A;Accession: I59509
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 832-909 <REX>
A;Cross-references: GB:I29395; NID:g459807; PIDN:AAA35809.1; PID:g459808
R;Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.
Mol. Cell. Biol. 7, 2597-2601, 1987
A;Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptiona
A;Reference number: I57622; MUID:87286898; PMID:3039351
A;Accession: I57622
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-191 <TAL>

insulin receptor p
insulin receptor p
protein-tyrosine k
insulin receptor -
insulin receptor-r
insulin receptor-r
insulin-like growt
insulin-like growt
insulin receptor -
insulin receptor -
protein-tyrosine k
protein-tyrosine k
tyrosine kinase Mp
mouse developmenta
protein-tyrosine k
protein-tyrosine k

A;Cross-references: GB:M16792; NID:g183983; PIDN:AAA58637.1; PID:g553332
C;Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
C;Genetics:
A;Gene: GDB:ERBB2; NGL; NEU; HER-2
A;Cross-references: GDB:120613; OMIM:164870
A;Map position: 17q21.1-17q21.1
A;Introns: 25/1; 75/3; 147/1; 883/3
A;Note: the list of introns is incomplete
C;Function:
A;Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
inase
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>
F;22-653/Domain: extracellular #status predicted <EXT>
F;70-304/Domain: EGF receptor extracellular domain repeat <EB1>
F;395-605/Domain: EGF receptor extracellular domain repeat <EB2>
F;654-675/Domain: transmembrane #status predicted <TM>
F;676-1255/Domain: intracellular #status predicted <INT>
F;718-983/Domain: protein kinase homology <KIN>
F;726-734/Region: protein kinase ATP-binding motif
F;68,124,187,259,530,571,629/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F;753/Active site: Lys #status predicted
F;1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 98.3%; Score 6694; DB 1; Length 1255;
Best Local Similarity 98.3%; Pred. No. 5,46-266;
Matches 1237; Conservative 5; Mismatches 9; Indels 8; Gaps 2;

Qy 1 MELAALCRWGLLALLPPGAASSTOVCTFNNFTVSFWLRVVKVSASHLE----LYQGCOVV 56
Db 1 MELAALCRWGLLALLPPGAASSTOVCTGTDK----LRLPASPETHLDMLRHLYLQGCOVV 56

Qy 57 QGNLELYPTNASLSFLQDIEQVGVYLAHNOVROVPLQRLRIVRGTLQFEDNYALAV 116
Db 57 QGNLELYPTNASLSFLQDIEQVGVYLAHNOVROVPLQRLRIVRGTLQFEDNYALAV 116

Qy 117 LDNGDPLNTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHNK 176
Db 117 LDNGDPLNTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHNK 176

Qy 177 NQALTATIDNRSRACHPCSPCKGSCWCGESSDCOSLRTVTCAGGCARCKGLPTDCC 236
Db 177 NQALTATIDNRSRACHPCSPCKGSCWCGESSDCOSLRTVTCAGGCARCKGLPTDCC 236

Qy 237 HEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCV 296
Db 237 HEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCV 296

Qy 297 TACPYNVLTSDVGSCVLPVLPVLPVLPVLPVLPVLPVLPVLPVLPVLPVLPVLPVLPV 356
Db 297 TACPYNVLTSDVGSCVLPVLPVLPVLPVLPVLPVLPVLPVLPVLPVLPVLPVLPV 356

Qy 357 TSANIQFAGCKKIFGSLAFIPESFDGDPASNTAPLOEQLOVETLEETGYLYISAMP 416
Db 357 TSANIQFAGCKKIFGSLAFIPESFDGDPASNTAPLOEQLOVETLEETGYLYISAMP 416

Qy 417 DSLPDLVSFQNLQVIRGRILHNGAYSLLTQGLGTSWGLRLSRLGSLALIHNTLHLCF 476
Db 417 DSLPDLVSFQNLQVIRGRILHNGAYSLLTQGLGTSWGLRLSRLGSLALIHNTLHLCF 476

Qy 477 VHTVPVQDLFRNPHOALLHTANRDEBCVGBGLACHQLCARGHCWGPPTQCVNCSQFLR 536
Db 477 VHTVPVQDLFRNPHOALLHTANRDEBCVGBGLACHQLCARGHCWGPPTQCVNCSQFLR 536

Qy 537 GOECVEECRVLOGLPREVYNARHCLPCHPECPONGSVTCGPEADOCVACHYKDPBPC 596
Db 537 GOECVEECRVLOGLPREVYNARHCLPCHPECPONGSVTCGPEADOCVACHYKDPBPC 596

Qy 597 VARCPGSKVDPDLSYMP1WKFPDEBGAQPCPCINCTHSCVDLDDKGCQPAEQASPLTSIIS 656
Db 1 MELAALCRWGLLALLPPGAASSTOVCTFNNFTVSFWLRVVKVSASHLE----LYQGCOVV 56
Db 1 MELAALCRWGLLALLPPGAASSTOVCTGTDK----LRLPASPETHLDIVRHLYLQGCOVV 56

Db 597 VARCPGSKVDPDLSYMP1WKFPDEBGAQPCPCINCTHSCVDLDDKGCQPAEQASPLTSIIS 656
Qy 657 AVVGILLVVVLGVVFGILIKRROOKIRKYTMRLLOETELVEPLTPSGAMPNQAQRILK 716
Db 657 AVVGILLVVVLGVVFGILIKRROOKIRKYTMRLLOETELVEPLTPSGAMPNQAQRILK 716
Qy 717 ETELKRVKVLGSGAFGTGVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAG 776
Db 717 ETELKRVKVLGSGAFGTGVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAG 776
Qy 777 VGSPPYVRLGICLTSTVQLVLTOLMPYGCGLLDHVRNRRGLSGQDLLNWCQIAKMSYL 836
Db 777 VGSPPYVRLGICLTSTVQLVLTOLMPYGCGLLDHVRNRRGLSGQDLLNWCQIAKMSYL 836
Qy 837 EDVRLVHRDLAARNVLKSPNHVKITDGLARLLDDIDETEHADGGKVP1KMMALSESILR 896
Db 837 EDVRLVHRDLAARNVLKSPNHVKITDGLARLLDDIDETEHADGGKVP1KMMALSESILR 896
Qy 897 RRTTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMV 956
Db 897 RRTTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMV 956
Qy 957 KCMIDSECRPRPRELVSFERNARDPQRPVVTIONEDLGPASPLDSTFYRSLLDDEMDGD 1016
Db 957 KCMIDSECRPRPRELVSFERNARDPQRPVVTIONEDLGPASPLDSTFYRSLLDDEMDGD 1016
Qy 1017 LVDAEYLVPOQGFCCPDPAAGAGVHRRHSSSTRSGGDLTLGLEPSEEEAPRPLA 1076
Db 1017 LVDAEYLVPOQGFCCPDPAAGAGVHRRHSSSTRSGGDLTLGLEPSEEEAPRPLA 1076
Qy 1077 PSEGAGSDVFDGDLGMAAKGLOSLTHDPSPQRYSEDPVPLPSETDGYVAPLTCSPQ 1136
Db 1077 PSEGAGSDVFDGDLGMAAKGLOSLTHDPSPQRYSEDPVPLPSETDGYVAPLTCSPQ 1136
Qy 1137 PEYVNPQVVRPQPPSPREGPLPAARPAAGATLERAKTSLPGKNGVWKFVAFGGAIVENPEY 1196
Db 1137 PEYVNPQVVRPQPPSPREGPLPAARPAAGATLERAKTSLPGKNGVWKFVAFGGAIVENPEY 1196
Qy 1197 LTPQGAAPQPPHPPAFSPAFDNLYWDDPPRGAAPPSTFKGTPTAENPEYLGLOVPV 1255
Db 1197 LTPQGAAPQPPHPPAFSPAFDNLYWDDPPRGAAPPSTFKGTPTAENPEYLGLOVPV 1255

RESULT 2
I48161
P-185 precursor - golden hamster
C;Species: Mesocricetus auratus (golden hamster)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999
C;Accession: I48161
R;Nakamura, T.; Ushijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishikawa
Gene 140, 251-255, 1994
A;Title: Cloning and activation of the Syrian hamster neu proto-oncogene.
A;Reference number: I48161; MID:94193007; PMID:7908275
A;Accession: I48161
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-1254 <RES>
A;Cross-references: GB:D16295; NID:G493236; PIDN:BA03801.1; PID:g747595
C;Genetics:
C;Gene: neu
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP
C;Region: protein kinase ATP-binding motif
F;726-734/Region: protein kinase ATP-binding motif

Query Match 86.3%; Score 5879.5; DB 2; Length 1254;
Best Local Similarity 86.2%; Pred. No. 8,9e-233;
Matches 1085; Conservative 60; Mismatches 105; Indels 9; Gaps 3;

Qy 1 MELAALCRWGLLALLPPGAASSTOVCTFNNFTVSFWLRVVKVSASHLE----LYQGCOVV 56
Db 1 MELAALCRWGLLALLPPGAASSTOVCTGTDK----LRLPASPETHLDIVRHLYLQGCOVV 56


```
Qy 57 QGNLELYLPNASLSFLQDIQEVQGVVLIHNNQVRQVPLQRLRIVRGTLFEDNYALAV 116
Db 57 QGNLELYLPANATLSFLQDIQEVQGYMLIAHSQVRHVPLQRLRIVRGTLFEDKYALAV 116
Qy 117 LDNGDPLNNTTPVTGASPGGLRELQLRLSLTEILKGGVLIQHNPOLCYQDTILWKDIFHN 176
Db 117 LDNRDPLNNTTATGRTPEGLRELQLRLSLTEILKGGVLIQHNPOLCYQDTILWKDIFRN 176
Qy 177 NQLALTLIDNRSRACHPCSPMKGSRGWESSEDCQSLTRTVCAAGCARCKGPLPTDCC 236
Db 177 NQLAPVDIDNRSRACPPCAPCKDNHCWASPEDCQTLTGCTIAPRAVPAARARLPTDCC 236
Qy 237 HEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMNPPEGRTYTFGASCV 296
Db 237 HEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMNPPEGRTYTFGASCV 296
Qy 297 TACPNYLSLTDVGSCTLVCPLNQEVTAEDGTQRCCKSPCARVCYGLGWEHLREVRV 356
Db 297 TTCPNYLSLTVGSCTLVCPLNQEVTAEDGTQRCCKSPCARVCYGLGWEHLRGARAI 356
Qy 357 TSANIQEFAGCKKIFGSLAPFESFDGDPASNTAPLOEQLVQFETLEEITGYLIYSAMP 416
Db 357 TSANIQEFAGCKKIFGSLAPFESFDGPNSSGIAPLTPQEQVQFETLEEITGYLIYSAMP 416
Qy 417 DSLPDLVSFQNLQVIRGRIHNGAYSITLQIGISWGLSLRLSRLGSLGALIHNTLCLF 476
Db 417 DSLHDLVSFQNLQVIRGRLVHDGAYSIALQIGIRWGLSLRLSRLGSLGLVLIHRNTLCLF 476
Qy 477 VHTVPDQLPRNPHQALLHTANRPEDECVCBGLACHOLCARGHCWGPGTQVCNCSOFLR 536
Db 477 VHTVPDQLPRNPHQALLHNSPSEBCKLDFACYPLCAHGHGWGPGTQVCNCSHFLR 536
Qy 537 GQECVEBCRVLQGLPREYVNAHCLPCHPEQCPQNGSVTCFPGPADQCVACAHYKDPFFC 596
Db 537 GQECVCKEVRWGLPREYVNGKCHLCPHPEQCPQNSTETCTGSEADQCTACPHYKDSFFC 596
Qy 597 VARCPGVKPDLSWMTWKPDDEGACQPCPINCTHSCVDLDDKGCQPAERASPLTISVS 656
Db 597 VARCPGVKPDLSWMTWKPDDEGMQCPCPINCTHSCVDLDERGCAERASPATSIIA 656
Qy 657 AVGILLAVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQOMRILK 716
Db 657 TVVGILLVFLVGVVVGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQOMRILK 716
Qy 717 ETELKRVKVLGSGAFGYKGIWIPDGENVKIPVAIKVLRNTPSPKANKELLDAYVMAG 776
Db 717 ETELKRVKVLGSGAFGYKGIWIPDGENVKIPVAIKVLRNTPSPKANKELLDAYVMAG 776
Qy 777 VGSPIVSRLLGICLTSTVQLVTQLMYPYGCLLDQVRENRGLSGQDLNWCQIAKGSYL 836
Db 777 LGSPIVSRLLGICLTSTVQLVTQLMYPYGCLLDQVRENRGLSGQDLNWCQIAKGSYL 836
Qy 837 EDVRLVHRDLAARNVLKSNHVKITDGLARLLDIDETEYHADGGKVPKWMALLESILR 896
Db 837 EDVRLVHRDLAARNVLKSNHVKITDGLARLLDIDETEYHADGGKVPKWMALLESILR 896
Qy 897 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMV 956
Db 897 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMV 956
Qy 957 KCMWIDSECRPRPRELVSFSESRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDMDMD 1016
Db 957 KCMWIDSECRPRPRELVSFSESRMARDPQRFVVIQNEDLGPSSPLDSTFYRSLLEDMDMD 1016
Qy 1017 LVDAEEYLVQOQGFCDPPAPGACGMVHRRSSSTRSGGDLTLGLPESEEEAPRSPILA 1076
Db 1017 LVDAEEYLVQOQGFCDPPAPGAGSTARRHRRSSSTRSGGELTLGNEPSEEPSPRSPILA 1076
Qy 1077 PSEGAGSDVFDGDLGMAAGLQSLPTHDPSPLOQYSEDPTVPLPSETDGVVAPLTCSPQ 1136
Db 1077 PSEGAGSDVFEGLGMAATKGPQISIPRDLSPLOQYSEDPTPLPSETDGVVAPLACSPQ 1136
Qy 1137 PEYVNPQDVRPQPPSPREGPLPAARAPAGATLERAKTILSPGKNGVYVKOVFAFGGAVENPEY 1196
```

RESULT 3

TVRTNU

protein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 11-Jun-1999

C;Accession: A24562; A61204

R;Bargmann, C.I.; Hung, M.C.; Weinberg, R.A.

Nature 319, 226-230, 1986

A;Title: The neu oncogene encodes an epidermal growth factor receptor-related protein.

A;Reference number: A24562; MUID:86118662; PMID:3945311

A;Accession: A24562

A;Molecule type: mRNA

A;Residues: 1-1260 <BAR>

A;Cross-references: EMBL:X03362; NID:g56745; PIDN:CAA27059.1; PID:g56746

R;Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohen,

Carcinogenesis 12, 1975-1978, 1991

A;Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals no

2-thiazolyl]formamide or N-methyl-N-nitrosourea.

A;Reference number: A61204; MUID:92035293; PMID:1682063

A;Accession: A61204

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 637-663, 'V', 665-702 <NAS>

A;Note: authors translated the codon GCA for residue 25 as Val

C;Genetics:

A;Gene: neu

C;Superfamily: epidermal growth factor receptor; protein kinase homology

C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosph

F;1-19/Domain: signal sequence #status predicted <SIG>

F;20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>

F;658-680/Domain: transmembrane #status predicted <TMN>

F;723-988/Domain: protein kinase homology <KIN>

F;731-739/Region: protein kinase ATP-binding motif

F;71,191,463,535,576,634/Binding site: carboxydrate (Asn) (covalent) #status predicted

F;691/Binding site: phosphate (Thr) (covalent) #status predicted

F;758/Active site: Lys #status predicted

F;882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 86.3%; Score 5876; DB 1; Length 1260;

Best Local Similarity 86.2%; Pred. No. 1.2e-232;

Matches 1087; Conservative 54; Mismatches 110; Indels 10; Gaps 4;

Qy 1 MELAALCRWGLLLALLPPGAASTQVCTFNNFTVSFWLRVPKVSASHLE---LYQGQGVV 56

Db 4 MELAANCRWGLLLALLPPGIAGTQVCTGDMK---LRLPASPETHLDMRLHLYQGQGVV 59

Qy 57 QGNLELYLPNASLSFLQDIQEVQGVVLIHNNQVRQVPLQRLRIVRGTLFEDNYALAV 116

Db 60 QGNLELYFPANASLSFLQDIQEVQGVVLIHNNQVRQVPLQRLRIVRGTLFEDKYALAV 119

Qy 117 LDNGDPLNNTTPVT-GASPGGLRELQLRLSLTEILKGGVLIQHNPOLCYQDTILWKDIFHK 175

Db 120 LDNRDPLNNTTPVTGRTPEGLRELQLRLSLTEILKGGVLIQHNPOLCYQDMVLMKQVFRK 179

Qy 176 NQALTLIDNRSRACHPCSPMKGSRGWESSEDCQSLTRTVCAAGCARCKGPLPTDC 235

Db 180 NQALPVDIDNRSRACPPCAPCKDNHCWASPEDCQTLTGCTIAGTCGACRCKGRLPTDC 239

Qy 236 CHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMNPPEGRTYTFGASC 295

Db 240 CHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMNPPEGRTYTFGASC 299

Qy 296 VTACPNYLSLTDVGSCTLVCPLNQEVTAEDGTQRCCKSPCARVCYGLGWEHLREVRV 355

Db 300 VTTCPNYLSLTVGSCTLVCPNNQEVTAEDGTQRCCKSPCARVCYGLGWEHLRGARA 359

A;Cross-references: GB:M16792; NID:g183983; PIDN:AAA58637.1; PID:g553332
C;Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
C;Genetics:
A;Gene: GDB:ERBB2; NGL; NEU; HER-2
A;Map position: 17q21.1-17q21.1
A;Introns: 25/1; 75/3; 147/1; 883/3
A;Note: the list of introns is incomplete
C;Function:
A;Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
inase
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>
F;22-653/Domain: extracellular #status predicted <EXT>
F;70-304/Domain: EGF receptor extracellular domain repeat <EE1>
F;395-605/Domain: EGF receptor extracellular domain repeat <EE2>
F;654-675/Domain: transmembrane #status predicted <TM>
F;676-1255/Domain: intracellular #status predicted <INT>
F;718-983/Domain: protein kinase homology <KIN>
F;726-734/Region: protein kinase ATP-binding motif
F;68.124,187,259,530,571,629/Binding site: carboxylate (Asn) (covalent) #status predicted
F;686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F;753/Active site: Lys #status predicted
F;1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 97.4%; Score 6619; DB 1; Length 1255;
Best Local Similarity 96.9%; Pred. No. 1.8e-265;
Matches 1221; Conservative 6; Mismatches 17; Indels 16; Gaps 2;

Qy 1 MELAAALCRWGLLLALLPPGAASCTGCTDMKRLRPASPETHDMLRLHYQGCQVQGNL 60
Db 1 MELAAALCRWGLLLALLPPGAASCTGCTDMKRLRPASPETHDMLRLHYQGCQVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIEVQGYVLIHAHQVQVQLRIRVGRQTLFEDNYALAVLNG 120
Db 61 ELTYLPTNASLSFLQDIEVQGYVLIHAHQVQVQLRIRVGRQTLFEDNYALAVLNG 120
Qy 121 DPLNNTTPTVGTASPGGLRELQRLSLEILKGGVLIQRNPOLCYQDTILWKDIFHKNNOLA 180
Db 121 DPLNNTTPTVGTASPGGLRELQRLSLEILKGGVLIQRNPOLCYQDTILWKDIFHKNNOLA 180
Qy 181 LTLIDTNRSRACHPCSPCKSRGWESSEDCOSLTRVTCAGGCARCKGPIPTDCCHQC 240
Db 181 LTLIDTNRSRACHPCSPCKSRGWESSEDCOSLTRVTCAGGCARCKGPIPTDCCHQC 240
Qy 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
Qy 301 YNLTSTDVGSCTLVCLPHNQEVTAEDGTQRCCKSPCARVCYGLGMQYIKANSKFIGIT 360
Db 301 YNLTSTDVGSCTLVCLPHNQEVTAEDGTQRCCKSPCARVCYGLGMHLEVRVAVTSAN 360
Qy 361 ELEFAGCKKIFGSLAFLPESPDGASNTAPLOPSOLOVFTLEBITGYLYISAMPDSLP 420
Db 361 IQEFAGCKKIFGSLAFLPESPDGASNTAPLOPSOLOVFTLEBITGYLYISAMPDSLP 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSLSLTQLGLISWGLRLSRELGSGLALIHNNHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSLSLTQLGLISWGLRLSRELGSGLALIHNNHLCFVHTV 480
Qy 481 PWDQLFRPHQALLTANRPEDECYVGEGLACHOLCARGHCWGPPTQCVCNSQFIRGQEC 540
Db 481 PWDQLFRPHQALLTANRPEDECYVGEGLACHOLCARGHCWGPPTQCVCNSQFIRGQEC 540
Qy 541 VEECRVLOGLPREYNARHCLPCHPECOPOGNSVTCFGEADQCACAHYKDPPECVARC 600
Db 541 VEECRVLOGLPREYNARHCLPCHPECOPOGNSVTCFGEADQCACAHYKDPPECVARC 600
Qy 601 PSGVKPDLSPYMPKPFDEBEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTISI 660

RESULT 2

TVRTNU

protein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 11-Jun-1999

C;Accession: A24562; A61204

R;Bartmann, C.I.; Hung, M.C.; Weinberg, R.A.

Nature 319, 226-230, 1986

A;Title: The neu oncogene encodes an epidermal growth factor receptor-related protein.

A;Reference number: A24562; MUID:86118662; PMID:3945311

A;Accession: A24562

A;Molecule type: mRNA

A;Residues: 1-1260 <BAR>

A;Cross-references: EMBL:X03362; NID:g56745; PIDN:CAA27059.1; PID:g56746

R;Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohen,

Carcinogenesis 12, 1975-1978, 1991

A;Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals no m

2-thiazolyl]formamide or N-methyl-N-nitrosourea.

A;Reference number: A61204; MUID:92035293; PMID:1682063

A;Accession: A61204

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 637-663,'V',665-702 <MAS>

A;Note: authors translated the codon GCA for residue 25 as Val

C;Genetics:

A;Gene: neu

C;Superfamily: epidermal growth factor receptor; protein kinase homology

C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho

F;1-19/Domain: signal sequence #status predicted <SIG>

F;20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 09:10:23 ; Search time 22.179 Seconds
(without alignments)
5413.772 Million cell updates/sec

Title: SEQ4-632-652-12
Perfect score: 6796
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVVPV 1249
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 73.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6619	97.4	1255	1 A24571	protein-tyrosine k
2	5860	86.2	1260	1 TVRTWU	protein-tyrosine k
3	5833	85.8	1254	2 I48161	p-185 precursor -
4	3115	45.8	1210	1 GQHUE	epidermal growth f
5	3082	45.4	1210	2 A53183	epidermal growth f
6	3051.5	44.9	1223	1 TVCHLV	epidermal growth f
7	2925.5	43.0	1308	2 A47253	epidermal growth f
8	2635	38.8	1166	1 S06142	protein-tyrosine k
9	2398.5	35.3	1342	2 A36223	kinase-related tra
10	2297.5	33.8	1339	2 JC4387	epidermal growth f
11	1723.5	25.4	698	1 TVFVLV	protein-tyrosine k
12	1660	24.4	604	1 TVYUHV	protein-tyrosine k
13	1617.5	23.8	1330	1 GQFE	epidermal growth f
14	1604	23.6	544	2 S35745	protein-tyrosine k
15	1597	23.5	545	2 S00727	kinase-related tra
16	1580	23.2	540	2 B4F776	protein-tyrosine k
17	1578	23.2	540	1 TVFVVB	protein-tyrosine k
18	1509	22.2	644	2 A36325	epidermal growth f
19	1282	18.9	1323	2 E88257	protein let-23 (im
20	1282	18.9	1374	2 S70712	protein-tyrosine k
21	1190	17.5	1369	2 S70713	protein-tyrosine k
22	1148	16.9	1717	1 A45558	epidermal growth f
23	1126	16.6	527	2 A42032	epidermal growth f
24	975.5	14.4	843	2 A27131	epidermal growth f
25	806.5	11.9	346	2 S13807	protein-tyrosine k
26	754.5	11.1	311	2 S13808	protein-tyrosine k
27	711	10.5	1363	2 T43220	insulin-like growt
28	681	10.0	1382	1 INHUR	insulin receptor p
29	679	10.0	1607	2 T43212	insulin-like growt

RESULT 1

A24571

protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human
N;Alternate names: c-erb-B-2 protein precursor; kinase-related transforming protein erbB
C;Species: Homo sapiens (man)
C;Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999
C;Accession: A24571; A25491; A44188; B44188; I59509; I57622
R;Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T.;
Nature 319, 230-234, 1986
A;Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth f.
A;Reference number: A24571; MUID:86118663; PMID:3003577
A;Accession: A24571
A;Molecule type: mRNA
A;Residues: 1-1255 <YAM>
R;Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.
Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985
A;Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epiderm
A;Reference number: A25491; MUID:86016729; PMID:2995967
A;Accession: A25491
A;Molecule type: DNA
A;Residues: 737-1031 <SEM>
A;Cross-references: GB:M11767; NID:gi82163; PIDN:AAA35808.1; PID:9553282
R;Coussens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg, P
Science 230, 1132-1139, 1985
A;Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chromos
A;Reference number: A44188; MUID:86070181; PMID:2999974
A;Accession: A44188
A;Molecule type: DNA
A;Residues: 740-910 <COU1>
A;Cross-references: GB:M12036; NID:gi83988; PIDN:AAA35978.1; PID:gi83989
A;Accession: B44188
A;Molecule type: mRNA
A;Residues: 1-517, 'RALL', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>
A;Cross-references: GB:M11730; NID:gi83986
R;King, C.R.; Kraus, M.H.; Aaronson, S.A.
Science 229, 974-976, 1985
A;Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.
A;Reference number: I59509; MUID:85272597; PMID:2992089
A;Accession: I59509
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 832-909 <REX>
A;Cross-references: GB:I29395; NID:9459807; PIDN:AAA35809.1; PID:9459808
R;Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.
Mol. Cell. Biol. 7, 2597-2601, 1987
A;Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptional
A;Reference number: I57622; MUID:87286898; PMID:3039351
A;Accession: I57622
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-191 <TAL>

insulin receptor p
insulin receptor-r
insulin receptor p
protein-tyrosine k
insulin receptor-r
insulin-like growt
insulin receptor -
insulin-like growt
insulin receptor -
insulin receptor (k
protein-tyrosine k
protein-tyrosine k
protein-tyrosine k
protein-tyrosine k
protein-tyrosine k

```
QY 638 DDKCPAEQASPLTSIYSAVV-GILLVVVLGVVFGILIKRQOKIRKYTWRRLLQETEL 696
Db 58 GLEGCP---NGSKTPSIAAGVVGGLLVVVGVLGIGLYLRRR-HIVRRTLRRLQEREL 113
QY 697 VEPLTPSGAMPNQAMRILKETELRKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLR 756
Db 114 VEPLTPSGEAPNQAHLRILKETEPKVKVVLGFGAGFTYKGLWIPEGEKVTIPVAIKELR 173
QY 757 ENTSPKANKEILDEAYVWAGVGSPIVSRLLGICLTSTVQLTQMLPYCCLLDHVRENRR 816
Db 174 EATSPKANKEILDEAYVWASVNDPHVCRLLGICLTSTVQLITQMLPYCCLLDYIREHKDN 233
QY 817 LGSODLLANWCQIAKMSYLEDVRLVHRDLAARNVLKSPNHVKITDPLGLARLLDIDETE 876
Db 234 IGSQYLLNWCQIAKMWYLEERHLVHRDLAARNVLVKTPODVKITDPLGLAKQGADEKE 293
QY 877 YHADGGKVPICKMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLE 936
Db 294 YHAEKGKVPICKMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIIPASEISSVLE 353
QY 937 KGERLPQPPICTIDVYIMVWKWMDSECRPRFRELSEFSEMRARDPORFVVIQ-NEDLG 995
Db 354 KGERLPQPPICTIDVYIMVWKWMSDADSRPKFRELIAEFSKMRARDPPRYLVIQDERMH 413
QY 996 PASPLDSTFYRSLLEDMDGLVDAEYLVPOQGFCCPDAPGAGGVMVHRRHRSSTRSG 1055
Db 414 LPSPDTSKPYRTLMEEDMEDIVDAEYLVPHQGF-----NSPST--- 454
QY 1056 GGDLTGLLEPSEEEAPRSP-----APSEGAGSDVFDGDLGMAAKGLQLPTHDPSPQLQ 1110
Db 455 -----SRTPLLSLSATSNNNSATNCIDRNG-----H----- 481
QY 1111 RYSEDPTVPLPSETDGYVAPLTCSPQPEYVQPDVVRPQPPSPREGPLPAARPAGAT-LER 1169
Db 482 -----PVREDGFL-----PAPEYVQ--LMPKKPSTAMVQNOIYNYISLTAISK 523
QY 1170 AKTLPSPKNGVVKDVFAGGAVENPEYL 1197
Db 524 LPMDRYQN-----SHSTAVDNPEYL 544
```

Search completed: July 22, 2003, 09:10:12
Job time : 31.9062 secs

```
Db 709 MRHVNQYATGPGY-----CAASPPRSKITSANLD-----VNMFIITGAV 749
Qy 663 LVVLGVVFGI-LIKRRQOKIRKYT--NRRLLQETVELVEPLTPSGAMPNQAOQMRILKETE 719
Db 750 LVPTTICILCVTYIICRQKAKETVMTMALSGRDSEPLRPSNIGANLCKLRIVKDAE 809
Qy 720 LRKVKVLSGAFGTGKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVGS 779
Db 810 LRKGVLMGAFGRVYKGVWVPEGENVKIPVAIKELLSTGAESSEEFLEAYIMASEEH 869
Qy 780 PVSLELLGICLTSTVOLATQMLPYCILLDHYVRENRLGSLQDILLNWCQIAKMSYLEDV 839
Db 870 VNLKLLAVCSSQMLLTQMLPLCCLLDYVRNRRDKIGSKALLNWSTQIAKMSYLEEK 929
Qy 840 RLVRHDLAARNVLV--SPNHVKITDFGLARLLDIDETEHADGGKVPKIKWMALESILR 896
Db 930 RLVRHDLAARNVLVRLLAGEDH----DFGLAKLLSSDSNEYKAAGRKPIKWLALCEIRN 985
Qy 897 RFRTHQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPICTIDVYIMV 956
Db 986 RVFTSKSDVWAFGVTIWEILLTFGQRPHENIPAKDIPDLIEVLGLKLEQPEICSLDIYCTLL 1045
Qy 957 KWMIDSCRRPRFRELSEFMRDPORFVVIQNEIDL--PASPLDSTFVRSLLDD-- 1012
Db 1046 SCHWLDAAAMRPTFKOLTTPAFAEFARDPGRYLAILGDKFTRLPA-----YTSQDEKDLI 1098
Qy 1013 -DMGDLVDAEVLVPOQGFPCPDAPGAGGVHHRHRSSTRSGGDLTLGLEPSEEAR 1071
Db 1099 RKLAPTTGSAIAKAPDDYLQKALGPS-----HRTDCT-----DEMP 1137
Qy 1072 -----RSLAPSEAGSDVFDG---DLGMAAGLQSLTHDPSPLQYSEDPTVPLPS 1122
Db 1138 KLNRYCKDPNSKNSSTGDERDSSAREVGVGNLR-----LDLPV 1176
Qy 1123 ETDGYVAPLTCSPQPEYVNDVRPQPSREGPLPAARPAAGATLAKTLSPQKNGVVK 1182
Db 1177 DEDDYLMP-TCQPGNNNNNN--NPNQNNMAAVGVAAGYM----- 1214
Qy 1183 DVFAFGGAVENPEYL---TPQGAAPQPH-----PPAFSP-AFDN 1219
Db 1215 DLIGVPVSDNPEYLLNNAQTILGVGESPIPTQTIGIPVMGGPTMEVKVMPGSEPTSSDH 1274
Qy 1220 LYYWD 1224
Db 1275 EYND 1279

RESULT 14
S35745
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus
C:Species: avian erythroblastosis virus
C:Date: 03-Mar-1994 #sequence_revision 26-May-1995 #text_change 28-Feb-1997
C:Accession: S35745
R:Vennstrom, B.
submitted to the EMBL Data Library, March 1993
A:Reference number: S35743
A:Accession: S35745
A:Molecule type: DNA
A:Residues: 1-544 <VEN>
A:Cross-references: EMBL:X12707
C:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; oncogene, phosphotransferase; transforming protein; tyrosine-specific P
F:135-400/Domain: protein kinase homology <KIN>
F:143-151/Region: protein kinase ATP-binding motif
F:170/Active site: Lys #status predicted

Query Match 23.8%; Score 1617; DB 2; Length 544;
Best Local Similarity 54.3%; Pred. No. 1.5e-59;
Matches 341; Conservative 70; Mismatches 125; Indels 92; Gaps 15;
```

```
Qy 578 GPEADQCACAHYKDPFPCVAVRCPOYIKANSKFIGITELPDEGACQPCPINCTHSCVDL 637
Db 1 GP--DHCWKCAHFTDGPCHVKACPAVLGENDTL-VMKYADANAVCQLCHPNCCTRGCKGP 57
Qy 638 DDKCPAEQASPLTSIVSAVV-CILLVVLGVVFGILLIKRRQOKIRKYTWRRLLQOETEL 696
Db 58 GLEGCP---NGSKTPSTAAGVVGGLCLLVVVGIGLYLRRR-HIVRKRTLRLRLQOEREL 113
Qy 697 VEPLTPSGAMPNQAOQMRILKETELRKVKVLSGAGFTGKGIWIPDGENVKIPVAIKVLR 756
Db 114 VEPLTPSGEAPNQAHILIKETEFKKVKVLFGFGAFGTGKGIWIPDGENVKIPVAIKELR 173
Qy 757 ENTSPKANKEILDEAYVMAGVGSYVSRLLGICLTSTVOLATQMLPYCILLDHYVRENRL 816
Db 174 EATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVOLITQMLPYCILLDHYREHKDN 233
Qy 817 LGSQDILLNWCQIAKMSYLEDLVLRDLAARNVLKSPNHVKITDFGLARLLDIDETE 876
Db 234 IGSQYLLNWCQIAKGMNYLEERHMHVRLDLAARNVLKTPQHVKITDFGLAKOLGADEKE 293
Qy 877 YHADGGKVPKIKWMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLE 936
Db 294 YHAEKGKVPKIKWMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGPAPSEISSVLE 353
Qy 937 KGERLPQPPICTIDVYIMVKNMIDSECRPRFRELSEFMRDPORFVVIQ-NEIDL 995
Db 354 KGERLPQPPICTIDVYIMVKNMIDSECRPRFRELSEFMRDPORFVVIQ-NEIDL 995
Qy 996 PASPLDSTFVRSLLDDMGDLVDAEYLVPOQGFPCPDAPGAGGVHHRHRSSTRSG 1055
Db 414 LPSPTDSKFTYRLMEEDMEDIVDAEYLVPHQGF--NSPST--- 454
Qy 1056 GGDITLGLFSEEARPSPL-----APSEGAGSDVFDGDLGMAAGLQSLPTHDPSPQL 1110
Db 455 -----SRPLLSLSATSNNSATNCIDRNG--H----- 481
Qy 1111 RYSEDDTVPPLPSETDGYVAPLTCSPQPEYVNDVRPQPSREGPLPAARPAAGAT-LER 1169
Db 482 -----PVREDGFL-----PAPEYVQ--LMPKPSSTAMVQVQIYNSLTAISK 523
Qy 1170 AKTLSPQKNGVVDVFAFGGAVENPEYL 1197
Db 524 LPIDSRYN-----SHSTAVDNPEYL 544

RESULT 15
S00727
kinase-related transforming protein (erbB) (EC 2.7.1.1) - avian erythroblastosis virus
C:Species: avian erythroblastosis virus
C:Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 28-Feb-1997
C:Accession: S00727
R:Scotting, P.; Vennstrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.
Oncogene Res. 1, 265-278, 1987
A:Title: Common site of mutation in the erbB gene of avian erythroblastosis virus mutant
A:Reference number: S00727; MUID:88217326; PMID:2897102
A:Accession: S00727
A:Molecule type: DNA
A:Residues: 1-545 <SCO>
A:Cross-references: EMBL:X06943
C:Genetics:
C:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; phosphotransferase
F:135-400/Domain: protein kinase homology <KIN>
F:143-151/Region: protein kinase ATP-binding motif

Query Match 23.7%; Score 1610; DB 2; Length 545;
Best Local Similarity 54.3%; Pred. No. 2.8e-59;
Matches 341; Conservative 69; Mismatches 126; Indels 92; Gaps 15;
```

```
Qy 578 GPEADQCACAHYKDPFPCVAVRCPOYIKANSKFIGITELPDEGACQPCPINCTHSCVDL 637
Db 1 GP--DHCWKCAHFTDGPCHVKACPAVLGENDTL-VMKYADANAVCQLCHPNCCTRGCKGP 57
```

Query Match 24.8%; Score 1688; DB 1; Length 604;
Best Local Similarity 51.7%; Pred No. 2e-62;
Matches 357; Conservative 76; Mismatches 131; Indels 126; Gaps 16;

QY 587 CAHYKDPFCVACPOYIKANSKFTIGITELPEEGACQPCPINCTHSCVDLDDKGCPEAQ 646
DB 3 CAHFDGPHCVKACPAVGLGENDTL-VKRYADANAVCQLCHPCNCTRGCKGLEGCP--- 58

QY 647 RASPLTSIVSAW-GILLVVLGVVFGILIKRQOKIRKYMRLLOTELEVEPLTPSGA 705
DB 59 NGSKTPSIAAGVGLGLLVVGLGIGLYLRR-HIVKRLRLRLLORELEVEPLTPSGE 117

QY 706 MPNQMRILKTELKRVKVLGSGAFVYXGWIPIPDGENVKIPVAIKVLRNTPSPKANK 765
DB 118 APNQAHILIKETEFKKVVLGSGAFGIYKGLWPEGEKVIPVAIKELREANSPKANK 177

QY 766 BILDAYVMAGVSPVSRLLIGICLTSTVQLTQMPYGLLDHVRNRLGSGODLW 825
DB 178 BILDAYVMASVNDPHVCRLLIGICLTSTVQLTQMPYGLLDYIREHKDNIGSQYLLW 237

QY 826 CMQIAKMSYLEDVPLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGGKVP 885
DB 238 CVQIAKGMNLYEERLVRDLAARNVLKTPHVKITDFGLAKLGADEKVEYHAEKGVP 297

QY 886 IKWMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLP 945
DB 298 IKWMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIPIAREISSVLEKGERLP 357

QY 946 ICTIDVIMVKKWIMIDSECRPRFELVSEFSRMARDPQRFVIO-NEDLGASPLDSTP 1004
DB 358 ICTIDVIMVKKWIMIDADSRPKFRELIAEFSKWARDPPRYLVIQGERMHLPSPTSDF 417

QY 1005 YRSLEDDMDGLVDAEELVYPOQFFCPDPAPGAGGVHHRHRSSTSGGDLTLGLE 1064
DB 418 YRLTEEDMEDIVDAEELVYPOHFF-----NSPST----- 449

QY 1065 PSEEAAPSPL-----APSEAGSDVDFDGLGMAAGLQSLTPHDPQLQRYSDPTVP 1119
DB 450 -----SRTPLLSLSATSNNSATCID-----RNCQGHVPREDSFVQRYSSDPTGN 495

QY 1120 LPSET-DGYVAPLTCPOPEVYNQFVRPQPPSPREGPLPAARPAATLERAKTLSPGK 1177
DB 496 FLEESIDGDL-----PAPEYVQ--LMPKPESTAM----- 524

QY 1178 NGVVKDVF-----GGAVENPEYLTPOGGAAPQHPHPPAFSPAFD 1218
DB 525 --VQNOYVNFISLTAISKLPMSRYONSHSTAVDNPYL-----NTNQSPLAKTVFE 574

QY 1219 NLYYWDQPPPERGAPPSTFKGTPTAENPEY 1248
DB 575 SFPYIQSNHQ-----INLDNPDY 594

RESULT 13
GOFPE
epidermal growth factor receptor - fruit fly (Drosophila melanogaster)
N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
C:Species: Drosophila melanogaster
C>Date: 17-Mar-1987 #sequence revision 17-Mar-1987 #text_change 11-Jun-1999
R:Liivneh, E.; Glazer, L.; Segal, D.; Schlessinger, J.; Shilo, B.Z.
Cell 40, 599-607, 1985
A:Title: The Drosophila EGF receptor gene homolog: conservation of both hormone binding
A:Reference number: A00640; MUID:85124611; PMID:2982499
A:Accession: A00640
A:Molecule type: DNA
A:Residues: 1-1330<LIV>
R:Wadsworth, S.C.; Vincent III, W.S.; Bilodeau-Wentworth, D.
Nature 314, 178-180, 1985
A:Title: A Drosophila genomic sequence with homology to human epidermal growth factor re
A:Reference number: A38021; MUID:85137938; PMID:2983232
A:Accession: A38021

A:Molecule type: DNA
A:Residues: 'A',832-866,'V',868-943,'QTPSLVK'<WAD>
A:Cross-references: EMBL:X02293; NID:97922; PIDN:CAA26157.1; PID:9929565
C:Comment: This sequence is tentative because the introns have not been identified.
C:Genetics:
A:Gene: FlyBase:Egfr
A:Cross-references: FlyBase:FBgn0003731
A:Map position: 2 57F
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
F:1-732/Domain: extracellular #status predicted <EXT>
F:733-764/Domain: transmembrane #status predicted <TM>
F:765-1330/Domain: intracellular #status predicted <INT>
F:808-1072/Domain: protein kinase homology <KIN>
F:816-824/Region: protein kinase ATP-binding motif
F:122,300,324,363,518,688,695,700/Binding site: carbohydrate (Asn) (covalent) #status pr
F:774/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:843/Active site: Lys #status predicted
F:1181/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status predict

Query Match 24.0%; Score 1630.5; DB 1; Length 1330;
Best Local Similarity 29.7%; Pred No. 9.8e-60;
Matches 412; Conservative 178; Mismatches 412; Indels 383; Gaps 40;

QY 80 VQGVLIHAHQVQVQLRIRIVRGTLF-----EDNYALAVLDNGDPLNNTTPVTGASP 134
DB 38 ITNVIVIGLDLIPCTLSVRLQLIRGTLFSLSVVEEKVALFV-----TY 81

QY 135 GGLRELQRLSITEILKGVLIQRNPQLCYQDTILWKDIFHKNQALALTLIDNRSRACHP 194
DB 82 SKMYTLBIPDLRDVLNGQVGHNNYLCMRTIQWSEIVSNGTDAYNYDFTAPERCEPK 141

QY 195 CSPNCKSRGVSESSDQSLTRTVAGGCA--RCKGPLPTDCCHECAAGCTGPKHSDC 252
DB 142 CHESCTHG-CWGEQPKKQCFKUTCSPOCAGGRCYGPKECHLFCAGCTGTQKDC 200

QY 253 LACLFHNSGICELHCPALVTYNTDTPESMPNPGRYTFGASCVTACPYNYLSTDVGSCT 312
DB 201 IACKNFDEAVSKECPMRKYNTPTVLETPNPEGKAYGATCVKECP-GHLLRDNGACV 259

QY 313 LVCPLHQEVTAEDGTORCEKSPCARVCVGLGMQVIAKANSKFIGITEL-----EPAG 366
DB 260 RSCPQDKMDKGE-----CVPKNGPCPKTC-----PGVTVLHAGNIDSPN 300

QY 367 CKKIFGSLAFUPESFDG--DPASNTA-----PQPEOLOVFETLEETGYLYISAWPDS 418
DB 301 CTVIDGNIRILDDQTFSGFDVYATVNGPRIPIDPERREVSVFVKELTGLNTEGTPQ 360

QY 419 LPDLVSFONLQVIRGILHNGAY-SLTIQGLIGISWGLRLSLRELGSGLALHHNTHLCFV 477
DB 361 FRNLSPFNLETIHGRQLMESMFAALAIKSSLSYLEMRNLKQISSGSVWIQHNRDLGV 420

QY 478 HTVPWDQLFRNPQALLHTANPEDEC----- 504
DB 421 SNIRWPAIQKEPEQKVYVNNENLRADLCGKFLTILISVOHNIIMHIFAICREKWNHLLGSV 480

QY 505 ----- 504
DB 481 ORGRLLSWHGSPVYLQELQFQWHLHRLWLYIQVINSITODKSNEHQLTDACYSPSVPT 540

QY 505 -----VG 506
DB 541 SLTIERARYAQSLAMELEQITARSASMRHSKTLPAEGRQVPRWFLGVCASARAGIA 600

QY 507 EGLA-----CHQLCARGHCWGPGTQCVNCSQFLRQECVCECRVLOGLPREVV---N 556
DB 601 EPLAGRAVCRKCHPLCELTNYGHEQVCSKCTHYKREQCETEC-----PADHYTDBE 654

QY 557 ARHCLPCHPECPQNGSVTCFGEADQCVACAHYK-----DPPF-----CVARCPQY 603
DB 655 QRECFQRHPEC---NG---CTGPGADCKSCRNFKLPDANETGPTVNSTMNCSTKCPLE 708

QY 604 IK-ANSKFIGITELPDEEGACQPCPINCTHSCVDLDDKGCPEAQRASPLTSIVSAVVGIL 662

Db 755 TDHMLAVGSLDHAHVRLLGLCPGSSQLVTOYLPLGSLLDHVHQHRETLGQPLLNGV 814
Qy 828 QIAKMSYLEDVRLVHRDLAARNLVKSPNHVKITDFGLARLLDDIDETEHADGGKVPK 887
Db 815 QIAKMYLLEHSMVHRDLAARNLVKSPNQVADFGVADLLPDDQLHSEAKTIK 874
Qy 888 WMALESILRRRPTHQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPIC 947
Db 875 WMALESIHFGKYTHQSDVMSYGVTVWELMTFGAEPYAGLRLAEIPDLLEKGERLAQPOIC 934
Qy 948 TIDVYIMVWKWMDSECRPRRELVSFMRARDPQRFVIVQNDLGPASPLDSTFTYRS 1007
Db 935 TIDVYIMVWKWMDENIRPTFKELANEFTRMARDPRYLVIKRA--GEGTP--PAAPPS 991
Qy 1008 LLEDDMDGLVDAEYLVPOQGFCCPDAPGAGGMVHRHRSSTRSGGDLTLGLEPSE 1067
Db 992 VLTTEL-----QEALEFEL-----DLDLLEAE 1017
Qy 1068 E-----EAPRSLAPSEG-----AGSDVFDGLGMGAAGLQ 1099
Db 1018 EGLATSLGALSPLPTGLTRPRGSSQLSPSSGYMPMNQSSILGEACLDSAVLGGREQFSR 1077
Qy 1100 SLPHDPSPLORYSEDPTVPLPSETDGYV-----APL-----SPQPE-----Y 1139
Db 1078 FISLH-PIPRGR-----PASESEGHVTGSEAELOEKVSVCRSRSRSPRPRGDSAY 1129
Qy 1140 VNQPDVRPQPPSPREP-----LPAARPAGATLERAKTLSP-GKNGV----- 1181
Db 1130 HSQRSLTPTVPLSPGLEEDGNGYVMPDTHLRCASSREGTLSSVGLSSVLGTEED 1189
Qy 1182 KDVFAGGAVENPEYLTPOGGAAPQPPHP 1210
Db 1190 ED-----EYEVYMKRGRSP-PRPP 1209

RESULT 11
TVFVLV
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian leukosis virus
N:Contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase
C:Species: avian leukosis virus, ALV
C>Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 11-Jun-1999
C:Accession: B00643; A00643
R:Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1985
A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and processing
A:Reference number: A00643; MUID:85228222; PMID:2980784
A:Accession: B00643
A:Molecule type: mRNA
A:Residues: 1-698 <N1>
A:Cross-references: GB:M10066; GB:M13881; NID:G211749; PID:AAA48763.1; PID:G211750
A:Note: in Genbank entry CHKERBBF, release 109.0, the source is designated as Gallus gallus
C:Comment: This protein is synthesized as a gag-env-erbB protein.
C:Genetics:
A:Gene: gag-env-erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific protein kinase
F:1-6/Product: gag protein (fragment) #status predicted <GAG>
F:7-59/Product: env protein (fragment) #status predicted <ENV>
F:60-698/Product: protein-tyrosine kinase erbB #status predicted <ERB>
F:194-459/Domain: protein kinase homology <KIN>
F:202-210/Domain: protein kinase homology <KIN>
F:202-210/Domain: protein kinase ATP-binding motif
F:229/Active site: Lys #status predicted

Query Match 25.5%; Score 1736.5; DB 1; Length 698;
Best Local Similarity 51.7%; Pred. No. 2.4e-64;
Matches 370; Conservative 80; Mismatches 141; Indels 125; Gaps 18;

Qy 578 GPEADQCVACAHYKDPFCVACRQYTKANSKFIGITELPDEGACQPCPINCTHSCVDL 637
Db 60 GP--DHCKMCAHFDGHCVCACGAVLGENDTL-VNKYADANAVCOLCHNCNTRCGCKGP 116
Qy 638 DDKGCAPRAQSPITSIASVY-GILVVLGVVFGVILIKRQOKIRKYMRLLOTEL 696

Db 117 GLEGP---NGSKTPSIAAGVYVGGLLCLVAVVGLGILYLR--HIVRKRTRLRLQEREL 172
Qy 697 VEPLTPSCAMPNOAMRILKETELRKVKVLGSGAGFTVYKGIWIPDGNVKIPVAIKVL 756
Db 173 VEPLTPSGEAPNQAHLRIKETEFKVKVLGSGAGFTVYKGIWIPDGNVKIPVAIKEL 232
Qy 757 ENTSPKANKEILDEAYVMAGVSGPVYRLLGICLTSTVQLVTQMLPMPYCGCLLDHVRNRR 816
Db 233 EATSPKANKEILDEAYVMASVDNPHVCELLGICLTSTVQLITQMLPMPYCGCLLDYIREHKDN 292
Qy 817 LGSODLLNWCQIAKMSYLEDVRLVHRDLAARNLVKSPNHVKITDFGLARLLDDIDETE 876
Db 293 IGSYLLNWCQIAKMSYLEDVRLVHRDLAARNLVKSPNHVKITDFGLARLLDDIDETE 352
Qy 877 YHADGGKVPKMALESILRRRPTHQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLE 936
Db 353 YHADGGKVPKMALESILRRRPTHQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLE 412
Qy 937 KGERLPQPPICTIDVYIMVWKWMDSECRPRRELVSFMRARDPQRFVIVQ-NEDLG 995
Db 413 KGERLPQPPICTIDVYIMVWKWMDADSRPKFELAEFKMARDPRYLVIQODERMH 472
Qy 996 PASPLDSTFTYRSLEDDMDGLVDAEYLVPOQGFCCPDAPGAGGMVHRHRSSTRSG 1055
Db 473 LPSPTDSKFYRTLMEEDMEDIVDAEYLVPHQGF-----NSPST--- 513
Qy 1056 GGDLTGLSEEEAPRSL-----APSEGAGSDVFDGLGMGAAGLQSLPTHDPSPLO 1110
Db 514 -----SRTPLLSSLSATSNNSATNCID-----RNGQGHVPREDSPVQ 550
Qy 1111 RYSDPTVPLPSET--DGYVAPLTCSPQPEYVNOQDVRPPOPPSPREGPLPAARAGATLE 1168
Db 551 RYSDPTGNTLESIDGFL-----PAPEYVNO--LMPKKPS----- 585
Qy 1169 RAKTLSPGKNGVVKDVF-----AFGAVENPEYLTPOGGAAPQPPHPAF 1213
Db 586 -----TAMVQNIYNNISLTAIKLPMDSRYQNSHSTAVDNPEYL-----NTNQSPLA 633
Qy 1214 SPAPDNLVYNDQ-----DPPE-----RGAPPSTFKGTPTAENPEYLGLDVP 1254
Db 634 KTVFESSPYIQSGNHQINLDNPDYQODFLPNETKPNGLLKVPAAENPEYLRVAAP 689

RESULT 12
TVYUHH
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain H)
C:Species: avian erythroblastosis virus
C>Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 11-Jun-1999
C:Accession: A00644; A38022
R:Yamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.
Cell 35, 71-78, 1983
A:Title: The erbB gene of avian erythroblastosis virus is a member of the src gene family
A:Reference number: A00644; MUID:84026539; PMID:6313229
A:Accession: A00644
A:Molecule type: DNA
A:Residues: 1-604 <YAM>
A:Cross-references: GB:K01216; NID:G209676; PID:AAA42400.1; PID:G209678
R:Deubure, B.; Henry, C.; Benaisa, M.; Biserte, G.; Claverie, J.M.; Saule, S.; Martin, S.
Science 224, 1456-1459, 1984
A:Title: Sequencing the erbA gene of avian erythroblastosis virus reveals a new type of
A:Reference number: A38022; MUID:84223957; PMID:6328658
A:Accession: A38022
A:Molecule type: DNA
A:Residues: 1-28, 'W', 30-139, 'F', 141-145, 'V', 147-152 <DEB>
A:Cross-references: GB:K02006
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific protein kinase
F:130-395/Domain: protein kinase homology <KIN>
F:138-146/Region: protein kinase ATP-binding motif
F:165/Active site: Lys #status predicted

Db 411 VFSNLTIGGRSLYNRGFSLLIMKNLVNTSLGFRSLKEISAGRIYISANRQLCYVHSLNW 470
Qy 483 DOLFNPQALLHTA-NRPEDECVEGLEACHOLCARGHCWGPPTQCVNCSQFLRGQCV 541
Db 471 TKVLRGPTTEERLDIKHNRPRDCVAEGKVCDPLCSSGGCGWGPQCLSCRYSGGVCV 530
Qy 542 EBCRVLOGLPREYVNAHCLPCHPCQPNQSVCTGPEADOCVACAHYKDPFFCARCP 601
Db 531 THCNFLNEPREFAHEACFCHPCQPMWEGTATNGSGDTCACQAFRDGPHCVSSCP 590
Qy 602 QYIKANSKFIG----ITELPDEGACQPCINCTHSCVDLDKGCAPQRA-----SPLTS 653
Db 591 HGV-----LGAKGPIYKVPDQNECRPCHENCTQCKGPELQDCLGQTLVLIGKTHLTM 644
Qy 654 IVSAVVGILLVVVLGVGILIKRQOKIR-KYTWRRLLLOETELVEPLTPSGAMPNQAM 712
Db 645 ALTVIAG--LVIFVMLGGTFLYMRGRRIQNRAMRYLERGESIEPLDPS-EKANKVLA 701
Qy 713 RILKETELRKVKVLGSGAGFTVYKGIWIPDGNVKI PVAIKVLRNTSPKANKEILDEAY 772
Db 702 RIFKETELRKVKVLGSGVFTVHKGVWIPGESIKIPVCIKVIEDKSGRQSFQATDHL 761
Qy 773 VMAGVSPVYSLRLGICLTSTVQLVTLMPYCGLLDHRNRLGSLQDLNMCQIAKG 832
Db 762 AIGSLDHAHIVRLGLCPGSSQLVQYPLGLSLLDHRQHRGALGPQLLLNWGVQIAKG 821
Qy 833 MSYLEDVRLVHDLAARNVLKSPNHVKITDGLARLLDIDTEYHADGCKVPIKMALE 892
Db 822 MYTLEHGMVHRNLARNVLLKSPQVQVADFPGLPPDKQLLYSEAKTPIKMALE 881
Qy 893 SILRRFTHOSDWSVGVTVWELMTFGAKPYGIPAREIPDLEKGERLPQPICTIDVY 952
Db 882 SHIFGKYTHOSDWSVGVTVWELMTFGAPYAGRLAEVDPDLEKGERLAQOICTIDVY 941
Qy 953 MIMVKCMWIDSECRPRFRELVEFSRMDARDPQRFVVIQNEIDLGA-----SPLDSTFYRSL 1009
Db 942 MVMVKCMWIDENIRPTEKELANEFTMRADPPRYLVIKRES-GPGIAPGPEPHGLTNKL 1000
Qy 1010 EDDMDGLVDAREYLVQGFPCPDPAFCAGGMVHRRHSSSTRSGGDLTLGLSP-SEE 1068
Db 1001 EBEVEPELDLDLEAEED-----NLATTILGSALSPLFVGTLANRP 1041
Qy 1069 EAPRSLAPSEAGSDVFDGLCMGAAGLQLSLPTH-DPSLQRYSEDPVPLP----- 1121
Db 1042 RQSQSLSPSSGY-NPMQNLGESCQESAVSGSSERCPVSLH-----PMPRGCLAS 1094
Qy 1122 SETDGYVA-----PLTCSQPE-----YVNPQDVRPQPPSPREGP----- 1156
Db 1095 ESSEGHVTGSEAELOEKVSMCRSRSRSPRPRGDSAYHSQRHSLTPTVTLSPGLEEE 1154
Qy 1157 -----LPAARPAGATLERAKTILSP-GKNGV-----KDVAFGGAVENPEYLTPOGA 1203
Db 1155 DVNGYVMPDTHLKGTPSSREGTLSSVGLSSVLGTEBEED-----BEYEMNRRRH 1206
Qy 1204 APOPHPPAPAFSPAFDNLNYWD-----QDPPERGAPPSTFKGTPTAENPEYL 1249
Db 1207 SP-PHPPRPSSLEELGYEYWDVGDVSDLSASLGSTQSCFLHPVIMPATGATPDEDVEYM 1263

RESULT 10
JC4387

epidermal growth factor receptor homolog precursor - rat

N/Alternate names: ErbB3 protein, HER3 protein

C/Species: Rattus norvegicus (Norway rat)

C/Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998

C/Accession: JC4387

R/Helliver, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.

Gene 165, 279-284, 1995

A/Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protein.

A/Reference number: JC4387; MUID:96096535; PMID:8522190

A/Accession: JC4387

A/Molecule type: mRNA

A/Residues: 1-1339 <HEL>

A/Cross-references: GB:U29339; NID:g915389; PID:g915390
A/Experimental source: liver
A/Note: The authors translated the codon AAC for residue 369 as Thr and GTT for residue
C/Comment: This protein is a functional heregulin receptor that transduces signals to th
C/Genetics:
A/Gene: ErbB3
C/Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo
C/Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane protein
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1339/Product: epidermal growth factor homolog #status predicted <MAT>
F:640-659/Domain: transmembrane #status predicted <TM>
F:705-970/Domain: protein kinase homology <KIN>
F:713-721/Region: protein kinase ATP-binding motif
F:939,1051,1156,1194,1196,1219,1257,1259,1273,1286,1325/Binding site: phosphate (Tyr) (

Query Match 33.9%; Score 2304.5; DB 2; Length 1339;

Best Local Similarity 40.4%; Pred. No. 2.4e-82;

Matches 521; Conservative 169; Mismatches 43; Indels 167; Gaps 36;

Qy 3 LAALCRWGLLIALLPAA---STOVCTGTDMLKRLPASPEHLDMLRLHLYQGCVVQGN 59
Db 7 LQVLC----FLLSARGSEMNSQAVCEPTLNGLSVTGDADNQYQTLYKLYKECEVMGN 62
Qy 60 LELTYLPTNASLSFLQDIQYQGVYLLIAHNOVROVPLQRLRIVRGTOQLFEDNYALAVLDN 119
Db 63 LEIVLTGHNADLSFLQWIREVTAVLVAMNEFSLPLNLRVVRGTQVYDGKFAIFVM-- 120
Qy 120 GDPLNNTPTVGTGSPGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNQL 179
Db 121 ---LNYNT---NSHALRQLKFTQLTEILSGGVYIEKNDKLCHMDTIDMRDVRVR-- 170
Qy 180 ALTIIDNRSRACHPCSPMCKSGCWSESSDDCSLRTVTCAGGC-ARCKPLPTDCCHE 238
Db 171 GAEIVVRKNGANCPCPEHVKCG-RWGGPDPDQILIKTICAPQNGECFPFNQCCHD 229
Qy 239 QCAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTPESMNPBGRVTFGASCVTA 298
Db 230 ECAGSGSPQDTCFACRRFNDSGACVPRCPDEPLVYNKLTQLEFPNPHTKYQYGGVCVAS 289
Qy 299 CPYNYLSTDVSCSTLVCPLHNOEVTAEQTCRCKSPCARVCYGL--GMQYIKANSKF 356
Db 290 CPHNFV-VDQTFVRAACPPDKMEVD-KHGLKQVCPGGLCPKACEGTGSGSGRYQTVDSN 347
Qy 357 IGITELEFAGCKIFGSLAFLESFDGDPASNTAPLOPEQLOVPEETEEITGYLYISAWP 416
Db 348 ID----GFVNCTKILGNLDLITGLNVDPWHKI PALDPEKLVNFTVREITGYLNIQSWP 403
Qy 417 DSLPDLVSFQNLQVIRGRIHNGAYS-LTLOGLGISWLGRLSRLSGLALIHHTHLC 475
Db 404 PHMNFVSFSLTITGRSLYNRGFSLLIMKNLVNTSLGFRSLKEISAGRVVISANQQLC 463
Qy 476 FVHTVPWDLFRNPHQALLHTA-NRPEDECVEGLEACHOLCARGHCWGPPTQCVNCSQF 534
Db 464 YHSLNTRLLRGSSEERLDIKYDRPLGECIAEGKVCDPCLCSSGGCGWPAQGCUSCRNY 523
Qy 535 LRGOECVBECKVLOGLPREYVNAHCLPCHPCQPNQSVCTGPEADOCVACAHYKDP 594
Db 524 SREGVCVTHCFLOGEPRFVHEAQCFSCHPECLPMEGTSTYNGSGSDACARCAHFRDGP 583
Qy 595 FCVARCPQYIIKANSKFIG----ITELPDEGACQPCINCTHSC--VDLDDKGCAPQRA 648
Db 584 HCVNSCPHGI-----LGAKGPIYKVPDAQNECRPCHENCTQCGNGPELQDCLGQAEVLM 637
Qy 649 SPLTSIVSAVVGILLVVVLGVVFGILLIKRQOKIR-KYTWRRLLLOETELVEPLTPSGAMP 707
Db 638 SKPHIVIAVTVG--LAVILMILGGSFLYMRGRRIQNRAMRYLERGESIEPLDPS-EKA 694
Qy 708 NQAOVRILKETELRKVKVLGSGAGFTVYKGIWIPDGNVKI PVAIKVLRNTSPKANKEI 767
Db 695 NKVLARIKFEELRKVKVLGSGVFTVHKGVWIPGESIKIPVCIKVIEDKSGRQSFQAV 754
Qy 768 LDEAYVMAGVSPVYSLRLGICLTSTVQLVTLMPYCGLLDHRNRLGSLQDLNMCQ 827

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:11:38 ; Search time 37.7958 Seconds
(without alignments)
4403.399 Million cell updates/sec

Title: SEQ4-5-25-12
Perfect score: 6774
Sequence: 1 MELAALCRWGLLIALLPPGA.....TFKGTPTAENPEYLGIDVPV 1249

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

```

Database :
1: A_Geneseq_101002.*
1: /SID$2/cgcdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID$2/cgcdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID$2/cgcdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID$2/cgcdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID$2/cgcdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID$2/cgcdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID$2/cgcdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID$2/cgcdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID$2/cgcdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID$2/cgcdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID$2/cgcdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID$2/cgcdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID$2/cgcdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID$2/cgcdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID$2/cgcdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID$2/cgcdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID$2/cgcdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID$2/cgcdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID$2/cgcdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID$2/cgcdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID$2/cgcdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID$2/cgcdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID$2/cgcdata/geneseq/geneseq-emb1/AA2002.DAT.*

```

SUMMARY

Result No.	Query Score	Match Length	ID	Description
1	6678	98.6	21 AAY92620	Human heregulin 2
2	6678	98.6	22 AAE12130	Human tyrosine kin
3	6678	98.6	22 AAB60167	HER2 transgene pla
4	6678	98.6	22 AAW74545	Human HER2 (FrbB2)
5	6672	98.5	17 AAU01111	HER-2/neu protein.
6	6672	98.5	20 AAW92406	Human HER-2/neu on
7	6672	98.5	21 AAB21198	Human HER-2/neu pr
8	6672	98.5	21 AAY94780	Amino acid sequenc
9	6672	98.5	22 AAB85458	Human HER-2/neu pr
10	6672	98.5	22 AAG898267	HER2/neu amino aci

[illegible]

FT /label= Cysteine_rich_domain
 FT 210..224
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 250..264
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 324..483
 FT /label= Ligand_binding_domain
 FT 325..339
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 369..383
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 465..479
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 484..623
 FT /label= Cysteine_rich_domain
 FT 579..593
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 624..654
 FT /label= Transmembrane_domain
 FT 632..652
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 653..667
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 655..1010
 FT /label= Tyrosine_kinase_domain
 FT 661..675
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 695..709
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 710..730
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 1011..1235
 FT /label= C-terminal_domain

WO200020027-A2.

13-APR-2000.

05-OCT-1999; 99WO-DK00525.

05-OCT-1998; 98DK-0001261.

20-OCT-1998; 98US-0105011.

(MEBI-) M & E BIOTECH AS.

Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I;
 Gautam A, Birk P, Karlsson G;

WPI; 2000-349917/30.

N-PSDB; AAA09455.

Inducing immune responses to weakly immunogenic, tumor associated
 peptide antigens for the treatment of breast and prostate cancer

Claim 62; Page 193-198; 220pp; English.

This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of
 Her2 can be used in the claimed method as an autovaccine to induce a CTL
 response. Subdominant CTL epitopes, antibody binding regions and
 cysteine residues involved in disulfide bonds are preserved in the
 immunogenized forms. Regions suitable for the insertion of foreign T
 helper epitopes were identified (see features table). The method

CC is used for inducing immune responses against weakly immunogenic
 CC cell-associated peptide antigens (PA) such as those associated with
 CC cancers (self-proteins), e.g. human prostate specific membrane antigen
 CC (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).
 CC The method comprises effecting simultaneous presentation by antigen
 CC producing cells (APCs) of the animals immune system of: (1) at least 1
 CC CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1
 CC B-cell group derived from the cell-associated PA; and (2) at least 1
 CC first T helper cell group which is foreign to the animal. Analogues of
 CC human PSM, human Her2 and human/murine FGF8b comprising a substantial
 CC part of all known and predicted CTL and B-cell epitopes of the respective
 CC PA and including at least one foreign T helper epitope are also claimed.
 CC The method is used to treat prostate, prostate/breast or breast cancer
 CC when the PA is human PSM, FGF8b and Her2, respectively.

XX SQ Sequence 1255 AA;

Query Match 98.6%; Score 6678; DB 21; Length 1255;

Best Local Similarity 98.3%; Pred. No. 0;

Matches 1234; Conservative 4; Mismatches 11; Indels 6; Gaps 1;

QY 1 MELAALCRWGLLLALLPFGAASQTCT-----QVIKANSKFIGITELLYQSCQVQGNL 54
 DB 1 MELAALCRWGLLLALLPFGAASQTCTGTDMKRLPASPETHLDMRLHLYQSCQVQGNL 60
 QY 55 ELTYLPTNASLSFLQDIOEVQGYVLIAHNOVRQVPLQRLIRIVRGTLQDPEDNYALAVLDNG 114
 DB 61 ELTYLPTNASLSFLQDIOEVQGYVLIAHNOVRQVPLQRLIRIVRGTLQDPEDNYALAVLDNG 120
 QY 115 DPLNTTPTVTGASPGGLRELQRLSTEILKGGVLIQRPQLCYQDTILMKDIFHKNNQLA 174
 DB 121 DPLNTTPTVTGASPGGLRELQRLSTEILKGGVLIQRPQLCYQDTILMKDIFHKNNQLA 180
 QY 175 LTLIDTNSRACHPCSPMKSGSRGWGESSEDCQSLTRIVCAGGCARCKGPLTDCHEQC 234
 DB 181 LTLIDTNSRACHPCSPMKSGSRGWGESSEDCQSLTRIVCAGGCARCKGPLTDCHEQC 240
 QY 235 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDFESMPNPEGRVTFGASCVTACP 294
 DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDFESMPNPEGRVTFGASCVTACP 300
 QY 295 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRAVTSAN 354
 DB 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRAVTSAN 360
 QY 355 IOEFAGCKKIFGSLAFPLPESFDGDPASNTAPLOPQLQVFTLEITGYLIYSANPDSLP 414
 DB 361 IOEFAGCKKIFGSLAFPLPESFDGDPASNTAPLOPQLQVFTLEITGYLIYSANPDSLP 420
 QY 415 DLSVFQNLQVIRGRILHNGAYSLTLQGLIGISWLGRLSRELGSGLALHNNTHLCFVHTV 474
 DB 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLIGISWLGRLSRELGSGLALHNNTHLCFVHTV 480
 QY 475 PWDQLFRNPQALLHTANRPEDECYGEGLACHQLCARGHCWGPPTQCVCNCSQFLRGQBC 534
 DB 481 PWDQLFRNPQALLHTANRPEDECYGEGLACHQLCARGHCWGPPTQCVCNCSQFLRGQBC 540
 QY 535 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFQPEADQCACAHYKDPDPPFCVARC 594
 DB 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFQPEADQCACAHYKDPDPPFCVARC 600
 QY 595 PSGVKPDLISYMPIWKFPDEEGACQPCINCHTSCVDLDDKGCAPASORASPLTSIVSAVVG 654
 DB 601 PSGVKPDLISYMPIWKFPDEEGACQPCINCHTSCVDLDDKGCAPASORASPLTSIVSAVVG 660
 QY 655 ILLVVVLGVFGILIKRQKIRKYTMERLLQETELVEPLTPSGAMPNQAMRIKRETEL 714
 DB 661 ILLVVVLGVFGILIKRQKIRKYTMERLLQETELVEPLTPSGAMPNQAMRIKRETEL 720
 QY 715 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEIIDEAIVMAGVGSPP 774
 DB 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEIIDEAIVMAGVGSPP 780

QY 775 YVSRLLGICLTSTVQLTQMPYGCCLLDHVRENRLGSGDILLNWCQTAKGMSYLEDVR 834
Db 781 YVSRLLGICLTSTVQLTQMPYGCCLLDHVRENRLGSGDILLNWCQTAKGMSYLEDVR 840
QY 835 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKKNMALESILRRRT 894
Db 841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKKNMALESILRRRT 900
QY 895 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVKWM 954
Db 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVKWM 960
QY 955 IDSECPRELYSESRWARDQRFVVIQNEIDLGASPLDSTFYRSLDEDDMGDLVA 1014
Db 961 IDSECPRELYSESRWARDQRFVVIQNEIDLGASPLDSTFYRSLDEDDMGDLVA 1020
QY 1015 EYLVVPOQGFCCPDPAFGAGMVHHRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1074
Db 1021 EYLVVPOQGFCCPDPAFGAGMVHHRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
QY 1075 AGSDVFDGLGMAAGLQSLPHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1134
Db 1081 AGSDVFDGLGMAAGLQSLPHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
QY 1135 NQDVRPQPPSPREGPLPAARPAGATLERAKTILSPGKNGVVKDVPFAGGAVENPEYLTQ 1194
Db 1141 NQDVRPQPPSPREGPLPAARPAGATLERAKTILSPGKNGVVKDVPFAGGAVENPEYLTQ 1200
QY 1195 GGAAPQHPPPAPFAPFDNLVYWDQDPPERGAPPSTFKGPTAENPEYLGLOVVP 1249
Db 1201 GGAAPQHPPPAPFAPFDNLVYWDQDPPERGAPPSTFKGPTAENPEYLGLOVVP 1255

RESULT 2
AAE12130
ID AAE12130 standard; Protein; 1255 AA.
XX
AC AAE12130;
XX
XX
DT 18-DEC-2001 (first entry)
XX
DE Human tyrosine kinase-type receptor, HER-2.
XX
KW Therapeutic compound; major histocompatibility complex; vaccine;
KW antigenic peptide; MHC; immunoregulatory; immune response; HER-2;
KW adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;
KW antigen presenting cell; human; tyrosine kinase-type receptor.
XX
OS Homo sapiens.
XX
FH Key
FT Region
FT 774..782
FT /note= "Antigenic epitope"
XX
PN WO200168677-A2.
XX
XX
PD 20-SEP-2001.
XX
XX 16-MAR-2001; 2001WO-US40328.
XX
XX 16-MAR-2000; 2000US-0527487.
PR
XX (GENZ) GENZYME CORP.
XX
XX Nicolette CA;
PI
XX
XX WPI; 2001-616284/71.
DR N-PSDB; AAD19731.
XX
XX
PT Novel synthetic therapeutic compound for inducing immune response and
PT for use in adoptive immunotherapy, has enhanced binding to major
PT histocompatibility molecules and enhanced immunoregulatory properties

XX Claim 4; Page 63-67; 69pp; English.
XX
CC The invention relates to synthetic therapeutic compounds (antigenic
CC peptides) with enhanced binding to major histocompatibility complex
CC (MHC) molecules and enhanced immunoregulatory properties relative
CC to their natural counterparts. Compounds of the invention are useful
CC for inducing an immune response in a subject and for use in adoptive
CC immunotherapy. They are useful as components of anti-cancer vaccines
CC and to expand immune effector cells that are specific for cancers
CC characterised by expression of the breast cancer antigen, HER-2.
CC Polynucleotides that encode peptides of the invention are useful as
CC hybridisation probes and as primers for the detection of genes of gene
CC transcripts that are expressed in antigen presenting cells (APCs), to
CC confirm transduction of polynucleotides into host cells. The present
CC sequence is human tyrosine kinase-type receptor, HER-2. Compounds
CC of the invention are designed based on the HER-2 antigenic peptide
CC (774-782).
XX
SQ Sequence 1255 AA;
Query Match 98.6%; Score 6678; DB 22; Length 1255;
Best Local Similarity 98.3%; Pred No. 0;
Matches 1234; Conservative 4; Mismatches 11; Indels 6; Gaps 1;
QY 1 MELAALCRWGLLLALLPPGAASSTOVCT-----QYIKANSKFIGITELYQGCQVVQGNL 54
Db 1 MELAALCRWGLLLALLPPGAASSTOVCTGTDMKRLRLPASPETHLDMLRLHLYQGCQVVQGNL 60
QY 55 ELTYLPTNASISFLQDIOEVQVYVLI AHNVQVQVPLQRLIRVGTQQLFEDNYALAVLDNG 114
Db 61 ELTYLPTNASISFLQDIOEVQVYVLI AHNVQVQVPLQRLIRVGTQQLFEDNYALAVLDNG 120
QY 115 DPLNNTTPTVGASPGGLRELQRLSLEILKGGVLIQRNPOLCYQDTILWKDIFHKNNOLA 174
Db 121 DPLNNTTPTVGASPGGLRELQRLSLEILKGGVLIQRNPOLCYQDTILWKDIFHKNNOLA 180
QY 175 LTLIDNRSRACHPCSPMKSGSRWGSESDCQSLTRTVACGGCARKGKPLTDCHEQC 234
Db 181 LTLIDNRSRACHPCSPMKSGSRWGSESDCQSLTRTVACGGCARKGKPLTDCHEQC 240
QY 235 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 294
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 295 YNYLSTDVGSCTLVCPHNOEVTADGTQRCCKSPCARVCYGLGMEHLREVRAVTSAN 354
Db 301 YNYLSTDVGSCTLVCPHNOEVTADGTQRCCKSPCARVCYGLGMEHLREVRAVTSAN 360
QY 355 IOEFAGCKKIFGSLAFPLPESFDGDPASNTAPLQPEQLQVFFLEITGYLTSAMPDSLP 414
Db 361 IOEFAGCKKIFGSLAFPLPESFDGDPASNTAPLQPEQLQVFFLEITGYLTSAMPDSLP 420
QY 415 DLSVFQNLQVIRGRILHNGAYSLTLQIGISWGLRSLRELGSGLALIHNNTHLCFVHTV 474
Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQIGISWGLRSLRELGSGLALIHNNTHLCFVHTV 480
QY 475 PWDQLFRPHQALLHTANRPEDECYGEGLACHOLCARGHCWGPGTQCNCQFIRGQBC 534
Db 481 PWDQLFRPHQALLHTANRPEDECYGEGLACHOLCARGHCWGPGTQCNCQFIRGQBC 540
QY 535 VEECRVLQGLPREYVYNARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 594
Db 541 VEECRVLQGLPREYVYNARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600
QY 595 PSGVKPDLSYMPIWKFPDEBAGACPCPINCTHSCVDLDDKGPAPQBRASPLTSIVSAVVG 654
Db 601 PSGVKPDLSYMPIWKFPDEBAGACPCPINCTHSCVDLDDKGPAPQBRASPLTSIVSAVVG 660
QY 655 ILLVVVLGVVFGILIKRRQOKIRKYTMRLLOQETELVEPLTPSGAMPNQAOQMRILKETEL 714
Db 661 ILLVVVLGVVFGILIKRRQOKIRKYTMRLLOQETELVEPLTPSGAMPNQAOQMRILKETEL 720

QY 715 RKVKVLSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAGVSP 774
Db 721 RKVKVLSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAGVSP 780
QY 775 YVSRLLGICLTSTVOLVTLQMPYGCCLLDHVRENRHGLGSODLLNWCQIAKMSYLEDVR 834
Db 781 YVSRLLGICLTSTVOLVTLQMPYGCCLLDHVRENRHGLGSODLLNWCQIAKMSYLEDVR 840
QY 835 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGGKVPKMWMALESILRRRT 894
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGGKVPKMWMALESILRRRT 900
QY 895 HQSDVMSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVKWM 954
Db 901 HQSDVMSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVKWM 960
QY 955 IDSECPRELYSESRWARDQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDDMGDLVDA 1014
Db 961 IDSECPRELYSESRWARDQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDDMGDLVDA 1020
QY 1015 EYLVVPOQGFCDPAPAGAGWVHRRHRSSTRSGGDLTLGLEPSEERAPRSLAPSEG 1074
Db 1021 EYLVVPOQGFCDPAPAGAGWVHRRHRSSTRSGGDLTLGLEPSEERAPRSLAPSEG 1080
QY 1075 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEYV 1134
Db 1081 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140
QY 1135 NQDVPREPSPREGPLPAARPAATLERAKTILSPKNGVVKDVFAFGAVENPEYLTPO 1194
Db 1141 NQDVPREPSPREGPLPAARPAATLERAKTILSPKNGVVKDVFAFGAVENPEYLTPO 1200
QY 1195 GGAAPQHPHPPAFDNLVYWDQPPERGAAPPSTFKGTPTAENPEYLGLDVVPV 1249
Db 1201 GGAAPQHPHPPAFDNLVYWDQPPERGAAPPSTFKGTPTAENPEYLGLDVVPV 1255

RESULT 3
AAB60167
ID AAB60167 standard; Protein; 1255 AA.
XX AAB60167;
AC AAB60167;
XX
DT 03-APR-2001 (first entry)
XX
DE HER2 transgene plasmid construct encoded protein.
XX
KW Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
XX antibody.
XX Homo sapiens.
OS Synthetic.
XX
XX WO200100244-A2.
XX
XX 04-JAN-2001.
XX
XX 23-JUN-2000; 2000WO-US17229.
XX
XX 25-JUN-1999; 99US-0141316.
XX 16-MAR-2000; 2000US-0189844.
XX
XX (GETH) GENENTECH INC.
XX
XX Erickson S, Schwall R;
XX WPI; 2001-061962/07.
XX DR N-PSDB; AAF24297.
XX
XX Treating tumors, particularly breast cancers, which overexpress an ErbB
XX receptor and does not respond to an anti-ErbB antibody, comprises
XX conjugating the antibody to a maytansinoid -
XX

PS Example 3; Fig 4; 92pp; English.
XX
CC The present invention provides a method of treating cancer by
CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In
CC particular, the antibody is directed against ErbB2 (also known as HER2
CC and p185neu). The method is particularly useful in the treatment of
CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.
XX
SQ Sequence 1255 AA;

Query Match 98.6%; Score 6678; DB 22; Length 1255;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1234; Conservative 4; Mismatches 11; Indels 6; Gaps 1;

QY 1 MELAALCRWGLLLALLPFGAASQTCTCTGTDMKRLPASPETHLDMRLHLYQGCVVQGNL 54
Db 1 MELAALCRWGLLLALLPFGAASQTCTGTDMKRLPASPETHLDMRLHLYQGCVVQGNL 60
QY 55 ELYLPTNASLFLQDIOEVQGYVLI AHNOVQVPLQRLRIVRGTQLPEDNVALAVLDNG 114
Db 61 ELYLPTNASLFLQDIOEVQGYVLI AHNOVQVPLQRLRIVRGTQLPEDNVALAVLDNG 120
QY 115 DFLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRPQLCYQDTILWKDIFHKNNQLA 174
Db 121 DFLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRPQLCYQDTILWKDIFHKNNQLA 180
QY 175 LTLIDTNRACHPCSPCKSGRSGESSEDCQSLTRTVACGGCARCKGPLTDCHEQC 234
Db 181 LTLIDTNRACHPCSPCKSGRSGESSEDCQSLTRTVACGGCARCKGPLTDCHEQC 240
QY 235 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 294
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 295 YNLSLTDVGSCTLVCLHNEQVTAEDGTQRCCKSPCARVCYGLGMBHLREVRVTSAN 354
Db 301 YNLSLTDVGSCTLVCLHNEQVTAEDGTQRCCKSPCARVCYGLGMBHLREVRVTSAN 360
QY 355 IOEPAGCKKIPGSLAFPLPESFDGDPASNTAPLQPEQLQVFTLEBITGYLIYSAMPDLSLP 414
Db 361 IOEPAGCKKIPGSLAFPLPESFDGDPASNTAPLQPEQLQVFTLEBITGYLIYSAMPDLSLP 420
QY 415 DLSVFQNLQVIRGRILHNGAYSLTLQGLIGISWGLSRLSRELGSGLALTHNTHLCFVHTV 474
Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLIGISWGLSRLSRELGSGLALTHNTHLCFVHTV 480
QY 475 PWDQLFRNPHQALLHTANRPEDECYGEGLACHQLCARGHCWGPGTQCVCNCSQFLRGQEC 534
Db 481 PWDQLFRNPHQALLHTANRPEDECYGEGLACHQLCARGHCWGPGTQCVCNCSQFLRGQEC 540
QY 535 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFPEADQCVACAHYKDPFPCVARC 594
Db 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFPEADQCVACAHYKDPFPCVARC 600
QY 595 PSGVKPDLSPYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCPCAPORASPLTSIVSAVVG 654
Db 601 PSGVKPDLSPYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCPCAPORASPLTSIVSAVVG 660
QY 655 ILLVVVLGVWFGLIKRRQKIRKYTMRLLEQETELVEPLTPSGAMPNQAQRILKETEL 714
Db 661 ILLVVVLGVWFGLIKRRQKIRKYTMRLLEQETELVEPLTPSGAMPNQAQRILKETEL 720
QY 715 RKVKVLSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAGVSP 774
Db 721 RKVKVLSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAGVSP 780
QY 775 YVSRLLGICLTSTVOLVTLQMPYGCCLLDHVRENRHGLGSODLLNWCQIAKMSYLEDVR 834
Db 781 YVSRLLGICLTSTVOLVTLQMPYGCCLLDHVRENRHGLGSODLLNWCQIAKMSYLEDVR 840
QY 835 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGGKVPKMWMALESILRRRT 894

Db 841 LVHRDLAARNLVKSPNHHVITDFGLARLLDIDETEHADGGKVPKIKWMALESILRRRT 900
 QY 895 HQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPPOPICTIDVYIMVKWM 954
 Db 901 HQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPPOPICTIDVYIMVKWM 960
 QY 955 ISECRPRELVSERMRARDQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDMGLVDA 1014
 Db 961 ISECRPRELVSERMRARDQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDMGLVDA 1020
 QY 1015 EEVLVPQQGFPCPDPAAGAGGVMHRRSSSTRSGGDLTLGLPSEEEAPRSLAPSG 1074
 Db 1021 EEVLVPQQGFPCPDPAAGAGGVMHRRSSSTRSGGDLTLGLPSEEEAPRSLAPSG 1080
 QY 1075 AGSDVFDGLGMAAGKLSLTHDPSPLQRYSEDTVPLPSETDGYVAPLTCSPQPEYV 1134
 Db 1081 AGSDVFDGLGMAAGKLSLTHDPSPLQRYSEDTVPLPSETDGYVAPLTCSPQPEYV 1140
 QY 1135 NQPDVVRPQPPSPREGPLPAARPAAGATLERAKTLSPKNGVVKDVPFAGGAVENPEYLTPO 1194
 Db 1141 NQPDVVRPQPPSPREGPLPAARPAAGATLERAKTLSPKNGVVKDVPFAGGAVENPEYLTPO 1200
 QY 1195 GGAAPQHPHPPAFSPAFDNLVYWDQDPPRGAPPSTFKGTPTAENPEYLGLDVVP 1249
 Db 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPRGAPPSTFKGTPTAENPEYLGLDVVP 1255
 RESULT 4
 AAU74545
 ID AAU74545 standard; Protein; 1255 AA.
 AC AAU74545;
 XX
 DT 23-APR-2002 (first entry)
 XX
 DE Human HER2 (ErbB2) polypeptide.
 XX
 KW Human; HER2; ErbB; epidermal growth factor receptor;
 KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
 KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
 KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
 KW glial disorder; astrocytal disorder; hypothalamic disorder;
 KW glandular disorder; macrophagal disorder; epithelial disorder;
 KW stromal disorder; blastocoealic disorder; inflammatory disorder;
 KW angiogenic disorder; immunological disorder.
 XX
 OS Homo sapiens.
 XX
 XX US2002001587-A1.
 PN
 XX
 XX 03-JAN-2002.
 XX
 PF 16-MAR-2001; 2001US-0811123.
 XX
 XX 16-MAR-2000; 2000US-189844P.
 PR
 PR 05-OCT-2000; 2000US-238327P.
 XX
 XX (ERIC/) ERICKSON S.
 PA (SCHW/) SCHWALL R.
 PA (SLIW/) SLIWKOWSKI M.
 XX
 XX Erickson S, Schwall R, Sliwowski M;
 PI
 XX WPI; 2002-163686/21.
 DR
 DR N-PSDB; ABK14058.
 XX
 XX Treating tumour characterised by overexpression of epidermal growth
 PT factor receptor, ErbB or cancer in mammal, comprises administering
 PT anti-ErbB antibody-maytansinoid conjugate to the mammal -
 XX
 XX Example 3; Fig 7; 93pp; English.
 PS
 XX The invention relates to treating a tumour in a mammal, where the tumour

CC is characterised by the overexpression of an epidermal growth factor
 CC receptor (ErbB) and does not respond or responds poorly, to treatment
 CC with an anti-ErbB antibody, comprising administering to the mammal an
 CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for
 CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
 CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
 CC prostate and bladder, preferably breast cancer. The breast cancer is a
 CC metastatic breast cancer or an aggressive form of metastatic breast
 CC cancer which overexpresses ErbB2. The method is also useful for treating
 CC neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal,
 CC epithelial, stromal, blastocoealic, inflammatory, angiogenic and
 CC immunological disorders. This sequence represents the human HER2 (ErbB2)
 CC polypeptide of the invention.
 XX
 SQ Sequence 1255 AA;
 Query Match 98.6%; Score 6678; DB 23; Length 1255;
 Best Local Similarity 98.3%; Pred. No. 0;
 Matches 1234; Conservative 4; Mismatches 11; Indels 6; Gaps 1;
 QY 1 MELAALCRWELLLALLPPGAASQTQVCT-----QYIKANSKFIGITELLYQCCVQVQNL 54
 Db 1 MELAALCRWELLLALLPPGAASQTQVCTGDMKRLPASPETHLDMRLHLYQCCVQVQNL 60
 QY 55 ELTYLPTNASLFLQDIOEVQGYVLIHNRQVPLQRLIRIVRGTLQDFEDNTALAVLDNG 114
 Db 61 ELTYLPTNASLFLQDIOEVQGYVLIHNRQVPLQRLIRIVRGTLQDFEDNTALAVLDNG 120
 QY 115 DPLNNTTPTVGTASPGGLRELQLRSLTEILKGGVLTORNPOLCYQDITLWKDIFHKNNOA 174
 Db 121 DPLNNTTPTVGTASPGGLRELQLRSLTEILKGGVLTORNPOLCYQDITLWKDIFHKNNOA 180
 QY 175 LTLIDTNRSRACHPCSPMKSRCSGSESDCQSLTRTVTCAGGCARCKGKPLTDCCHEQC 234
 Db 181 LTLIDTNRSRACHPCSPMKSRCSGSESDCQSLTRTVTCAGGCARCKGKPLTDCCHEQC 240
 QY 235 AAGCTGPKXSDCLACLFHNFHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 294
 Db 241 AAGCTGPKXSDCLACLFHNFHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
 QY 295 NYLSTDVSGCTLVCPHNOEVTAEADGTQCEKSKPCARVCYGLGMEHLREVRAVTSAN 354
 Db 301 NYLSTDVSGCTLVCPHNOEVTAEADGTQCEKSKPCARVCYGLGMEHLREVRAVTSAN 360
 QY 355 IOEFAGCKKI FGSFLAPLPESDGDPASNTAPLOPELOVFTELEITGYLISAMPDSLP 414
 Db 361 IOEFAGCKKI FGSFLAPLPESDGDPASNTAPLOPELOVFTELEITGYLISAMPDSLP 420
 QY 415 DLSVFQNLQVIRGRILHNGAYSLTLQGLIGISWLGRLSLRELGSGLALIHNTLHLCFVHTV 474
 Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLIGISWLGRLSLRELGSGLALIHNTLHLCFVHTV 480
 QY 475 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHOLCARGHCWGPPTQCVNCSQFLRGQBC 534
 Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHOLCARGHCWGPPTQCVNCSQFLRGQBC 540
 QY 535 VEECRVLOGLPREYVYVNRHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFPFCVARC 594
 Db 541 VEECRVLOGLPREYVYVNRHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFPFCVARC 600
 QY 595 PSGVRPDLISYMPIMKFPDEEGACQPCINCTHSCVDLDDKGCAPAQBRASPLTSIVSAVVG 654
 Db 601 PSGVRPDLISYMPIMKFPDEEGACQPCINCTHSCVDLDDKGCAPAQBRASPLTSIVSAVVG 660
 QY 655 ILLVVVLGVVFGILIKRRQOKIRKYMRLRLOETELVEPLTPSGAMPNQAQMRILKETEL 714
 Db 661 ILLVVVLGVVFGILIKRRQOKIRKYMRLRLOETELVEPLTPSGAMPNQAQMRILKETEL 720
 QY 715 RKVKVLGSGAGFVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEIIDEAYVMAGVQSP 774
 Db 721 RKVKVLGSGAGFVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEIIDEAYVMAGVQSP 780
 QY 775 YVSRLLIGICTSTVQLVTQLMPYGCLLDHVRENRRGLSGSQDILLNWCMIQAKGMSYLEDVYR 834

Db 781 YVSRLLGICLTSTVQLVTQMPYGCCLLDHVRENRLGSDLLNWCQIAGKMSYLEDVR 840
QY LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGGKVPKMWALESIILRRFT 894
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGGKVPKMWALESIILRRFT 900
QY HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYMWKCM 954
Db 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYMWKCM 960
QY IDSECPREPRELVSERWARDQRFVVIQNEDLGPASPLDSTFYRSLLDDDDMDLVA 1014
Db 961 IDSECPREPRELVSERWARDQRFVVIQNEDLGPASPLDSTFYRSLLDDDDMDLVA 1020
QY 1015 EYLVFQOQFFCDPAPGAGMWHRRSSSTSGGGDLTLGLEPSESEAPRSLAPSEG 1074
Db 1021 EYLVFQOQFFCDPAPGAGMWHRRSSSTSGGGDLTLGLEPSESEAPRSLAPSEG 1080
QY 1075 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1134
Db 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
QY 1135 NOPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVDVFAFGAVENPEYLTPO 1194
Db 1141 NOPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVDVFAFGAVENPEYLTPO 1200
QY 1195 GGAAPQHPHPPAPSPAFDNLVYWDQPPERGAPPSTFKGTPTAENPEYLGLDVVP 1249
Db 1201 GGAAPQHPHPPAPSPAFDNLVYWDQPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 5

AAW01111
ID AAW01111 standard; Protein; 1255 AA.

XX AC AAW01111;

DT 01-JAN-1997 (first entry)

XX DE HER-2/neu protein.

XX KW HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
KW breast cancer; ovary cancer; colon cancer; lung cancer;
KW prostate cancer; immunisation; tumour; vaccine; vector.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers
FT Domain 676..1255
FT /label= Intracellular domain
FT /note= "claimed domain, useful for immunisation"

XX PN W09630514-AL.

XX XX

XX PD 03-OCT-1996.

XX PF 28-MAR-1996; 96WO-US01689.

XX PR 31-MAR-1995; 95US-0414417.

XX XX (UNIW) UNIV WASHINGTON.

XX XX Cheever MA, Disis ML;

XX XX WPI; 1996-455361/45.

XX XX N-PSDB; AAT40739.

XX XX DNA encoding HER-2-neu poly:peptide(s) - used for prevention or
PT treatment of malignancies with which the HER-2/neu oncogene is
PT associated

XX XX Claim 2; Page 56-61; 71pp; English.

XX CC Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is
CC the product of the HER-2/neu oncogene (see also AAT40739). The
CC protein is over-expressed in various cancers, including breast,
CC ovarian, colon, lung and prostate. The intracellular domain of the
CC protein can be used to immunise an animal against a malignancy with
CC which the oncogene is associated. The polypeptide can be produced
CC in transformed host cells for use in immunisation. Alternatively,
CC animal cells are transfected in vivo or ex vivo with a viral vector
CC that directs expression of the polypeptide.

XX SQ Sequence 1255 AA;
Query Match 98.5%; Score 6672; DB 17; Length 1255;
Best Local Similarity 98.2%; Pred. No. 0;
Matches 132; Conservative 5; Mismatches 12; Indels 6; Gaps 1;

QY 1 MELAALCRWGLLLALLPFGAASQTCT-----QYIKANSKFIGITELYQCCVQVQNL 54
Db 1 MELAALCRWGLLLALLPFGAASQTCTGTDMLRLPASPETHLMLRLHYQCCVQVQNL 60
QY 55 ELTYLPTNASLSFLQDIOEVQYVLIHNOVROVPLQRLRIVRGTQLPEDNYALAVLDNG 114
Db 61 ELTYLPTNASLSFLQDIOEVQYVLIHNOVROVPLQRLRIVRGTQLPEDNYALAVLDNG 120
QY 115 DPLNNTTPTVGTASPGGLRELQRLSLEILKGGVLIQRPOLCYQDTILWKDIFHKNQOLA 174
Db 121 DPLNNTTPTVGTASPGGLRELQRLSLEILKGGVLIQRPOLCYQDTILWKDIFHKNQOLA 180
QY 175 LTLIDTNRSRACHPCSPCKSGRSGESSEDCOSLTRTVAGGCAKCGPLPTDCCHEQC 234
Db 181 LTLIDTNRSRACHPCSPCKSGRSGESSEDCOSLTRTVAGGCAKCGPLPTDCCHEQC 240
QY 235 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESNPNEGRTYTFGASCVTACP 294
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESNPNEGRTYTFGASCVTACP 300
QY 295 NYLSTDVGSCTLVCPHNOEVTADGTQCEKSKPCARVCYGLGMEHLREVRVTSAN 354
Db 301 NYLSTDVGSCTLVCPHNOEVTADGTQCEKSKPCARVCYGLGMEHLREVRVTSAN 360
QY 355 IOEFAGCKKIFGSLAFPLESFDGDPASNTAPLOPEQLQVFTELEITGYLISAMPDSL 414
Db 361 IOEFAGCKKIFGSLAFPLESFDGDPASNTAPLOPEQLQVFTELEITGYLISAMPDSL 420
QY 415 DLSVFQNLQVIRGRILHNGAYSLTLQGLISWLGRLSRLGSLALIHNTHLCFVHTV 474
Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLISWLGRLSRLGSLALIHNTHLCFVHTV 480
QY 475 PWDQLFRNPHQALLHTANRPEDECYGEGLACHQLCARGHCWGPPTQCVNCSQFLRGQBC 534
Db 481 PWDQLFRNPHQALLHTANRPEDECYGEGLACHQLCARGHCWGPPTQCVNCSQFLRGQBC 540
QY 535 VEECRVLQGLPREYVYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 594
Db 541 VEECRVLQGLPREYVYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600
QY 595 PSGVKPDLSPYMPIWKFPDEEGACQPCPINCSTHSCVDLDDKGPAPQASPLTSISAVVG 654
Db 601 PSGVKPDLSPYMPIWKFPDEEGACQPCPINCSTHSCVDLDDKGPAPQASPLTSISAVVG 660
QY 655 ILLVVVLGVFGILLIKRQOKIRKYTMRLRLQETELVEPLTPSGAMPNQAQRILKETEL 714
Db 661 ILLVVVLGVFGILLIKRQOKIRKYTMRLRLQETELVEPLTPSGAMPNQAQRILKETEL 720
QY 715 RKVKVLSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVSP 774
Db 721 RKVKVLSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVSP 780
QY 775 YVSRLLGICLTSTVQLVTQMPYGCCLLDHVRENRLGSDLLNWCQIAGKMSYLEDVR 834
Db 781 YVSRLLGICLTSTVQLVTQMPYGCCLLDHVRENRLGSDLLNWCQIAGKMSYLEDVR 840

QY 835 LVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKWKMALESILRRRT 894
DB 841 LVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKWKMALESILRRRT 900
QY 895 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPPICTIDVTMIMVKWM 954
DB 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPPICTIDVTMIMVKWM 960
QY 955 IDSECRPRFELVSEFMRARDPQREVVIQNEIDLGPASPLDSTFYRSLLEDDDMGLVDA 1014
DB 961 IDSECRPRFELVSEFMRARDPQREVVIQNEIDLGPASPLDSTFYRSLLEDDDMGLVDA 1020
QY 1015 EBYLVPOQGFPCDPAPAGGVMHRRSSSTRSGGDLTLGLEPSEEEAPRPLAPSG 1074
DB 1021 EBYLVPOQGFPCDPAPAGGVMHRRSSSTRSGGDLTLGLEPSEEEAPRPLAPSG 1080
QY 1075 AGSDVFDGLGMAAGLQSLPHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEV 1134
DB 1081 AGSDVFDGLGMAAGLQSLPHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEV 1140
QY 1135 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTLPSPGKNGVVDVFAFGGAVENPEYLTPO 1194
DB 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTLPSPGKNGVVDVFAFGGAVENPEYLTPO 1200
QY 1195 GGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1249
DB 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 6

AAW92406
ID AAW92406 standard; Protein; 1255 AA.
XX
AC AAW92406;
XX
DT 21-APR-1999 (first entry)
XX
DE Human HER-2/neu oncogene protein.
XX
KW HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;
KW malignancy; treatment; tumour.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Region 676..1255
FT /note= "region which elicits immune response"
XX
PN US5869445-A.
XX
PD 09-FEB-1999.
XX
PF 01-APR-1996; 96US-0625101.
XX
PR 01-APR-1996; 96US-0625101.
PR 17-MAR-1993; 93US-0033644.
PR 12-AUG-1993; 93US-0106112.
PR 31-MAR-1995; 95US-0414417.
XX
PA (UNIW) UNIV WASHINGTON.
XX
PI Cheever MA, Disis ML;
XX
DR WPI; 1999-152835/13.
DR N-PSDB; AAX01912.
XX
PT Use of HER-2/neu polypeptides - for eliciting an immune response to
PT an HER-2/neu associated malignancy, particularly for treating or
PT preventing tumours
XX
PS Claim 3; Column 31-38; 26pp; English.
XX
CC This sequence represents the human HER-2/neu oncogene protein. A fragment

CC of this protein is used in a method for eliciting or enhancing an immune
CC response to HER-2/neu protein. The polypeptide can stimulate T cells and
CC B cells to produce an immune response to the HER-2/neu protein. The
CC method can be used for immunisation against a malignancy in which the
CC HER-2/neu oncogene is associated and in the treatment of an existing
CC tumour, or to prevent tumour occurrence or reoccurrence.
XX
SQ Sequence 1255 AA;
Query Match 98.5%; Score 6672; DB 20; Length 1255;
Best Local Similarity 98.2%; Pred No. 0;
Matches 1232; Conservative 5; Mismatches 12; Indels 6; Gaps 1;
QY 1 MELAALCRWGLLLALLPPGAASTQVCT-----QYIKANSKFITELLYQCCVQVQNL 54
DB 1 MELAALCRWGLLLALLPPGAASTQVCTGTDMLRLPASPTHLDMLRLHYLQCCVQVQNL 60
QY 55 ELTYLPTNASLFLQDIOEVQGYVLIAHNOVRQVPLQRLIRVGTQLFEDNVALAVLNG 114
DB 61 ELTYLPTNASLFLQDIOEVQGYVLIAHNOVRQVPLQRLIRVGTQLFEDNVALAVLNG 120
QY 115 DPLNNTTPTVGASPGGLRELQRLSLEILKGGVLIQORNPOLCYQDTILWKDIFHKNNOLA 174
DB 121 DPLNNTTPTVGASPGGLRELQRLSLEILKGGVLIQORNPOLCYQDTILWKDIFHKNNOLA 180
QY 175 LTLIDTNRSRACHPCSPCKSGRSGESSEDQSLTRTVCAAGGCARCKGPLTDCCHEQC 234
DB 181 LTLIDTNRSRACHPCSPCKSGRSGESSEDQSLTRTVCAAGGCARCKGPLTDCCHEQC 240
QY 235 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRTFGASCVTACP 294
DB 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRTFGASCVTACP 300
QY 295 YNYLSTDVGSCTLVCPHNOEVTAEQTCRCKSKPCARVCYGLGMEHLREVRVTSAN 354
DB 301 YNYLSTDVGSCTLVCPHNOEVTAEQTCRCKSKPCARVCYGLGMEHLREVRVTSAN 360
QY 355 IQEFAAGCKKI FGLSLAFLPESFDGDPASNTAPLQPEQLQVTFLEBITGYLIYSAMPDSL 414
DB 361 IQEFAAGCKKI FGLSLAFLPESFDGDPASNTAPLQPEQLQVTFLEBITGYLIYSAMPDSL 420
QY 415 DLSVFQNLQVIRGRILHNGAYSLTLQGLIGISWLGRLSRLRELGLALIHNNTHLCFVHTV 474
DB 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLIGISWLGRLSRLRELGLALIHNNTHLCFVHTV 480
QY 475 PMDQLFRNPHOALLHTANRPEDECYEGGLACHOLCARGHCWGPQCVCNCSQFLRGQSC 534
DB 481 PMDQLFRNPHOALLHTANRPEDECYEGGLACHOLCARGHCWGPQCVCNCSQFLRGQSC 540
QY 535 VEECRVLQGLPREYVYNARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 594
DB 541 VEECRVLQGLPREYVYNARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600
QY 595 PSGVKPDLISYMPIWKFPDEEGACQPCINCHTSCVDLDDKGPAPASORASPLTSIVAVVG 654
DB 601 PSGVKPDLISYMPIWKFPDEEGACQPCINCHTSCVDLDDKGPAPASORASPLTSIVAVVG 660
QY 655 ILLVVVLGVFGILIKRQOKIRKYTMERLLQETELVEPLTPSGAMPNQAORILKETEL 714
DB 661 ILLVVVLGVFGILIKRQOKIRKYTMERLLQETELVEPLTPSGAMPNQAORILKETEL 720
QY 715 RKVKVLGSGAFTGVYKGIWIPDGENVKIPVAIKVRENTSPKANKEILDEAYVMAGVSGP 774
DB 721 RKVKVLGSGAFTGVYKGIWIPDGENVKIPVAIKVRENTSPKANKEILDEAYVMAGVSGP 780
QY 775 YVSRLLGICLSTVOLVTQMLPYGCLLDHVRNRLGSLQDILLNWCMIAGKMSYLEDVR 834
DB 781 YVSRLLGICLSTVOLVTQMLPYGCLLDHVRNRLGSLQDILLNWCMIAGKMSYLEDVR 840
QY 835 LVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKWKMALESILRRRT 894
DB 841 LVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKWKMALESILRRRT 900

QY 895 HQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPICITIDVYMWKCM 954
 DB 901 HQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPICITIDVYMWKCM 960
 QY 955 IDSECRPRFRELVSERSMARDPQRFVIONEDLGPASPLDSTFYRSLLDDDDMDGLVDA 1014
 DB 961 IDSECRPRFRELVSERSMARDPQRFVIONEDLGPASPLDSTFYRSLLDDDDMDGLVDA 1020
 QY 1015 EYLVFQOGFFCDPAPGAGGVMHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1074
 DB 1021 EYLVFQOGFFCDPAPGAGGVMHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
 QY 1075 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1134
 DB 1081 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
 QY 1135 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTLPQKNGVVDVFAFGGAVENPEYLTPO 1194
 DB 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERPKTLPQKNGVVDVFAFGGAVENPEYLTPO 1200
 QY 1195 GGAAPQHPHPPAFSPAFDNLVYWDQPPERGAPPSTFKGTPTAENPEYLGLDVVP 1249
 DB 1201 GGAAPQHPHPPAFSPAFDNLVYWDQPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 7

ID AAB21198 standard; protein; 1255 AA.

AC AAB21198;

DT 12-JAN-2001 (first entry)

DE Human HER-2/neu protein.

XX Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;
 KW breast cancer; prostate cancer; ovarian cancer; lung cancer;
 KW colon cancer.

XX Homo sapiens.

XX WO200044899-A1.

XX 03-AUG-2000.

XX 28-JAN-2000; 2000WO-US02164.

XX 29-JAN-1999; 99US-0117976.

XX (CORI-) CORIXA CORP.

XX (SMIK) SMITHKLINE BEECHAM.

XX Cheever MA, Gheyssen D;

XX WPI; 2000-505976/45.

XX N-PSDB; AAA89736.

PT HER-2/neu extracellular domain/phosphorylation domain fusion proteins
 PT useful for vaccinating against breast, ovarian, colon, lung and
 PT prostate cancers -

XX Claim 52; Fig 7; 128pp; English.

CC The present sequence is the human HER-2/neu protein. It is a member of
 CC the tyrosine kinase family of receptor-like glycoproteins and shows
 CC homology to the epidermal growth factor receptor (EGFR). It probably
 CC plays a part in cell growth and/or differentiation. The HER-2/neu
 CC gene is an oncogene. An HER-2/neu fusion protein comprising a
 CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
 CC domain may be used to treat or prevent cancer by eliciting or
 CC enhancing an immune response to the HER-2/neu protein. It may be used
 CC to treat malignancies such as breast, ovarian, colon, lung and
 CC prostate cancers, and may be used as an antigen to vaccinate against

CC these neoplasias.
 XX Sequence 1255 AA;
 SQ Query Match 98.5%; Score 6672; DB 21; Length 1255;
 Best Local Similarity 98.2%; Pred. No. 0;
 Matches 1232; Conservative 5; Mismatches 12; Indels 6; Gaps 1;
 QY 1 MELAALCRWGLLALLPPGAASQTCTCTGTDMKLRLPASPEHLDMLRLHYLQCCVQVQNL 54
 DB 1 MELAALCRWGLLALLPPGAASQTCTCTGTDMKLRLPASPEHLDMLRLHYLQCCVQVQNL 60
 QY 55 ELTYLPTNASLSFLQDIOEVQYVLI AHNVQVPLQRLRI VRGTQLPFDNVALAVLDNG 114
 DB 61 ELTYLPTNASLSFLQDIOEVQYVLI AHNVQVPLQRLRI VRGTQLPFDNVALAVLDNG 120
 QY 115 DFLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 174
 DB 121 DFLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
 QY 175 LTLIDTNSRACHPCSPMKSGRSGESSEDCQSLTRTVACGGCARGKGPLTDCHEQC 234
 DB 181 LTLIDTNSRACHPCSPMKSGRSGESSEDCQSLTRTVACGGCARGKGPLTDCHEQC 240
 QY 235 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 294
 DB 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 QY 295 YNYLSTDVGSCTLVCPLNHNEVTAEDGTQRCCKSPCARVCYGLGMEHLREVAVTSAN 354
 DB 301 YNYLSTDVGSCTLVCPLNHNEVTAEDGTQRCCKSPCARVCYGLGMEHLREVAVTSAN 360
 QY 355 IOEFAGCKKIFGSLAFIPESFDGDPASNTAPLOPQLOVFETLEETITGYLIYSANPDSL 414
 DB 361 IOEFAGCKKIFGSLAFIPESFDGDPASNTAPLOPQLOVFETLEETITGYLIYSANPDSL 420
 QY 415 DLSVFQNLQVTRGRLIHNGAYSLTLQGLIGSWLGRSLRELGSGLALIHNNHLCFVHTV 474
 DB 421 DLSVFQNLQVTRGRLIHNGAYSLTLQGLIGSWLGRSLRELGSGLALIHNNHLCFVHTV 480
 QY 475 PMDQLFRNPQALLHTANRPEDECYGEGLACHQLCARGHCGPPTQCVCNCSOFLRGQEC 534
 DB 481 PMDQLFRNPQALLHTANRPEDECYGEGLACHQLCARGHCGPPTQCVCNCSOFLRGQEC 540
 QY 535 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDPFPCVARC 594
 DB 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDPFPCVARC 600
 QY 595 PSGVKPDLSPYMPIWKFPDEEGACQPCPINCSTHSCVDLDDKGPAPORASPLTSIVSAVVG 654
 DB 601 PSGVKPDLSPYMPIWKFPDEEGACQPCPINCSTHSCVDLDDKGPAPORASPLTSIVSAVVG 660
 QY 655 ILLVVVLGVFGILIKRQOKIRKYTMRLLLQETELVEPLTPSGAMPNQAQMRILKETEL 714
 DB 661 ILLVVVLGVFGILIKRQOKIRKYTMRLLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
 QY 715 RKVKVLGSGAGTGYKGIWIPDGENVKIPVAIKVLRNTSPKANKIILDEATVMAVGQSP 774
 DB 721 RKVKVLGSGAGTGYKGIWIPDGENVKIPVAIKVLRNTSPKANKIILDEATVMAVGQSP 780
 QY 775 YVSRLLGICLTSTVOLVTQMLPYGCLLDHVRNRLGSLQDLNWCMIKAGMSYLEDVR 834
 DB 781 YVSRLLGICLTSTVOLVTQMLPYGCLLDHVRNRLGSLQDLNWCMIKAGMSYLEDVR 840
 QY 835 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPFKWMALESILRRRT 894
 DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPFKWMALESILRRRT 900
 QY 895 HQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPICITIDVYMWKCM 954
 DB 901 HQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPICITIDVYMWKCM 960
 QY 955 IDSECRPRFRELVSERSMARDPQRFVIONEDLGPASPLDSTFYRSLLDDDDMDGLVDA 1014

Db 961 IDSECRPRRELVSFESRWARDPQREVVIONEDLGASPLDSTFYRSLLDDMDGLVDA 1020
 QY 1015 EYLVPQOQFFCPDPAPGAGGMVHHRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1074
 Db 1021 EYLVPQOQFFCPDPAPGAGGMVHHRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
 QY 1075 AGSDVPDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1134
 Db 1081 AGSDVPDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
 QY 1135 NOPDVRPQPPSPREGPLPAARAGATLERAKTILSPCKNGVWVDVFAFGGAVENPEYLTPO 1194
 Db 1141 NOPDVRPQPPSPREGPLPAARAGATLERPKTILSPCKNGVWVDVFAFGGAVENPEYLTPO 1200
 QY 1195 GGAAPOPHPPAFSPAFDNLTYWDQPPPERGAPPSTFKGPTAENPEYLGLOVPV 1249
 Db 1201 GGAAPOPHPPAFSPAFDNLTYWDQPPPERGAPPSTFKGPTAENPEYLGLOVPV 1255

RESULT 8

AAY84780

ID AAY84780 standard; Protein; 1255 AA.

XX AC AAY84780;

XX DT 08-AUG-2000 (first entry)

XX DE Amino acid sequence of the SPLICE erbb-2 receptor protein.

XX KW SPLICE erbb-2 receptor protein; cell transformation disorder; cancer;

XX KW tumor cell proliferation; tissue degeneration; arthropathy;

XX KW bone resorption; inflammatory disease; degenerative disorder;

XX KW wound healing.

XX OS Homo sapiens.

XX PN WO2000020579-A1.

XX PD 13-APR-2000.

XX PF 01-OCT-1999; 99WO-CA00912.

XX PR 02-OCT-1998; 98US-0165192.

XX PA (UYMC-) UNIV MCMASTER.

XX PI Muller WJ, Siegel PM;

XX PS WPI; 2000-303768/26.

XX DR N-PSDB; AAA14812.

XX PT Nucleic acid encoding an erbb 2 receptor protein designated SPLICE

XX PT erbb-2, inhibitors of the protein are useful for treatment of cancer -

XX PS Claim 3; Fig 2; 60pp; English.

XX CC The present sequence represents a SPLICE erbb-2 receptor protein. The

XX CC protein has an in-frame deletion of 16 amino acids, 2 of which are

XX CC conserved cysteine residues, compared to the unspliced protein. The

XX CC erbb-2 polynucleotide is used to construct probes for detecting

XX CC disorders of cell transformation such as cancer. Antibodies to the

XX CC protein may be used to detect SPLICE erbb-2 in a sample. Agents

XX CC (e.g. antisense oligonucleotides) which inhibit the expression of

XX CC SPLICE erbb-2 are useful for reducing tumor cell proliferation and

XX CC treating cancer. Substances which stimulate SPLICE erbb-2 are useful

XX CC for treating conditions involving damaged cells including conditions

XX CC in which degeneration of tissue occurs, such as arthropathy, bone

XX CC resorption, inflammatory diseases, degenerative disorders of the

XX CC central nervous system and wound healing.

XX XX Sequence 1255 AA;

XX SQ

Query Match 98.5%; Score 6672; DB 21; Length 1255;
 Best Local Similarity 98.2%; Pred. No. 0;
 Matches 1232; Conservative 5; Mismatches 12; Indels 6; Gaps 1;
 QY 1 MELAALCRWGLLLALLPPCAASTVCT-----QYIKANSKFIGITELLYQCCVQVQNL 54
 Db 1 MELAALCRWGLLLALLPPCAASTVCTGTDMKLRLPASPETHLDMRLHLYQCCVQVQNL 60
 QY 55 ELTYLPTNASLSFLQDIOEVQYVLIHNRQVQLRQLRIVRGTLQFEDNYALAVLNG 114
 Db 61 ELTYLPTNASLSFLQDIOEVQYVLIHNRQVQLRQLRIVRGTLQFEDNYALAVLNG 120
 QY 115 DPLNNTTPTVGTASPGGLRELQRLSLEILKGGVLIQRNPOLCYQDILWKDIFHKQNOLA 174
 Db 121 DPLNNTTPTVGTASPGGLRELQRLSLEILKGGVLIQRNPOLCYQDILWKDIFHKQNOLA 180
 QY 175 LTLIDTNRSEACHPCSPCKSGRSGESSEDCQSUTRTVCAGGCARCKGPLTDCHEOC 234
 Db 181 LTLIDTNRSEACHPCSPCKSGRSGESSEDCQSUTRTVCAGGCARCKGPLTDCHEOC 240
 QY 235 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 294
 Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 300
 QY 295 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSPCARVCYGLGMEHLREVRVTSAN 354
 Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSPCARVCYGLGMEHLREVRVTSAN 360
 QY 355 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFEITGLYISAMPDSDUP 414
 Db 361 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFEITGLYISAMPDSDUP 420
 QY 415 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLRELGSGLALIHNNTHLCFVHTV 474
 Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLRELGSGLALIHNNTHLCFVHTV 480
 QY 475 PWDQLFRPHQALLHTANRPEDECYVGEGLACHQLCARGHCWGPGTQCVNCSQFIRGQBC 534
 Db 481 PWDQLFRPHQALLHTANRPEDECYVGEGLACHQLCARGHCWGPGTQCVNCSQFIRGQBC 540
 QY 535 VEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPPECVARC 594
 Db 541 VEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPPECVARC 600
 QY 595 PSGVKPDLSPYMPKFPDEEGACQPCINCTHSCVDLDDKGPAPQASPLTSIVSAVVG 654
 Db 601 PSGVKPDLSPYMPKFPDEEGACQPCINCTHSCVDLDDKGPAPQASPLTSIVSAVVG 660
 QY 655 ILLVVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 714
 Db 661 ILLVVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
 QY 715 RKVKVLGSGAGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVQSP 774
 Db 721 RKVKVLGSGAGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVQSP 780
 QY 775 YVSRLLGICLSTVOLVTQLMPYGLLDHVRNRRGLSGQDLLNWCMIKAGMSYLEVDV 834
 Db 781 YVSRLLGICLSTVOLVTQLMPYGLLDHVRNRRGLSGQDLLNWCMIKAGMSYLEVDV 840
 QY 835 LVHRDLAARNVLKSPNHNKVTDFGLARLLDIDETEHADGKVPFKWMALESILRRRPT 894
 Db 841 LVHRDLAARNVLKSPNHNKVTDFGLARLLDIDETEHADGKVPFKWMALESILRRRPT 900
 QY 895 HQSDVMSVGVVWELMTGAKPYDGIPIAREIPDLLEKGERLPPOPICITDVMIMVKWM 954
 Db 901 HQSDVMSVGVVWELMTGAKPYDGIPIAREIPDLLEKGERLPPOPICITDVMIMVKWM 960
 QY 955 IDSECRPRRELVSFESRWARDPQREVVIONEDLGASPLDSTFYRSLLDDMDGLVDA 1014
 Db 961 IDSECRPRRELVSFESRWARDPQREVVIONEDLGASPLDSTFYRSLLDDMDGLVDA 1020
 QY 1015 EYLVPQOQFFCPDPAPGAGGMVHHRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1074

Db 1021 EYLVPQGFPCDPAPGAGMWHRRSSSTRSGGDLTLGLEPSEEAAPSLAPSEG 1080
QY 1075 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVAPLTCSPQPEV 1134
Db 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVAPLTCSPQPEV 1140
QY 1135 NQDVRPQPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAGGAVENPEYLTQ 1194
Db 1141 NQDVRPQPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAGGAVENPEYLTQ 1200
QY 1195 GGAAPQHPHPPAPSPAFDNLVYWDQPPRGAPPSTFKGTPTAENPEYLGDLVPV 1249
Db 1201 GGAAPQHPHPPAPSPAFDNLVYWDQPPRGAPPSTFKGTPTAENPEYLGDLVPV 1255

RESULT 9

AAB85458

ID AAB85458 standard; Protein; 1255 AA.

XX AC AAB85458;

DT 25-SEP-2001 (first entry)

DE Human HER-2/neu protein.

XX Antigen-presenting cell; immunogenic; immune response; HER-2/neu;

KW oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.

XX OS Homo sapiens.

XX WO200153463-A2.

XX 26-JUL-2001.

XX 19-JAN-2001; 2001WO-US01850.

XX 21-JAN-2000; 2000US-0177545.

XX (CORI-) CORIXA CORP.

XX Cheever MA, Hand-Zimmermann S;

XX WPI; 2001-476112/51.

XX N-PSDB; AAH23392.

XX New antigen-presenting cells, useful as vaccines for eliciting or
PT enhancing an immune response to HER-2/neu protein, particularly useful
PT for treating or preventing cancer, e.g. breast cancer -

XX Claim 2; Page 41-46; 49pp; English.

XX The invention provides an isolated antigen-presenting cell, which
CC expresses at least an immunogenic portion of a polypeptide that produces
CC an immune response to HER-2/neu protein. The antigen-presenting cells are
CC useful as vaccines for eliciting or enhancing an immune response to
CC HER-2/neu protein, particularly in treating or preventing malignancies in
CC which the HER-2/neu oncogene is associated. Specifically, these are
CC useful for treating or preventing cancer, e.g. breast cancer, ovarian,
CC colon, lung or prostate cancers. The present sequence represents
CC the human HER-2/neu protein (also known as p185 or c-erbB2).

XX Sequence 1255 AA;

Query Match 98.5%; Score 6672; DB 22; Length 1255;

Best Local Similarity 98.2%; Pred. No. 0;

Matches 1232; Conservative 5; Mismatches 12; Indels 6; Gaps 1;

QY 1 MELAALCRWGLLALLPPGAASSTQVCT-----QYIKANSKFTGITITELLYQSCVVGQNL 54

Db 1 MELAALCRWGLLALLPPGAASSTQVCTGDMKRLRLPASPTHLDMLRHLYQSCVVGQNL 60

QY 55 ELTYLPTNASLSFLQDIQEVQGVLYIAHNVQVPLQRLRIVRGTLQFEDNYALAVLDNG 114

Db 61 ELTYLPTNASLSFLQDIQEVQGVLYIAHNVQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
QY 115 DLNNNTTPTVGASPGGLRELQRLSLEITELKGGVLIQRPOLCYQDTILWKDIFHKNNQLA 174
Db 121 DLNNNTTPTVGASPGGLRELQRLSLEITELKGGVLIQRPOLCYQDTILWKDIFHKNNQLA 180
QY 175 LTLIDITNRSRACHPCSPCKSGRCWGESSEDCQSLTRTVACGGCARGKGPLTDCCHEQC 234
Db 181 LTLIDITNRSRACHPCSPCKSGRCWGESSEDCQSLTRTVACGGCARGKGPLTDCCHEQC 240
QY 235 AAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 294
Db 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 295 YNYLSTDVGSCTLVCPLNHNVTAEDGTORCEKSKPCARVCYGLGMEHLRVRVTSAN 354
Db 301 YNYLSTDVGSCTLVCPLNHNVTAEDGTORCEKSKPCARVCYGLGMEHLRVRVTSAN 360
QY 355 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOVFETLEBITGYLIYSAMPDSL 414
Db 361 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOVFETLEBITGYLIYSAMPDSL 420
QY 415 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLGSLALIHNNTHLCFVHTV 474
Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLGSLALIHNNTHLCFVHTV 480
QY 475 PWDQLFRPHQALLHTANRPEDECYVGEGLACHOLCARGHCWGPQTQVCNCSQFIRGQBC 534
Db 481 PWDQLFRPHQALLHTANRPEDECYVGEGLACHOLCARGHCWGPQTQVCNCSQFIRGQBC 540
QY 535 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFPCVARC 594
Db 541 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFPCVARC 600
QY 595 PSGVKPDLSPYMPKPFDEGACQPCPINCTHSCVDLDDKGPAPQASPLTSIYSAVVG 654
Db 601 PSGVKPDLSPYMPKPFDEGACQPCPINCTHSCVDLDDKGPAPQASPLTSIYSAVVG 660
QY 655 ILLVVVLGVVFGILIKRRQOKIRKYMRLRLQETELVEPLTPSGAMPNQAQMRILKETEL 714
Db 661 ILLVVVLGVVFGILIKRRQOKIRKYMRLRLQETELVEPLTPSGAMPNQAQMRILKETEL 720
QY 715 RKVKVLGSAGFTVYKGIWIPDGENVKIPVAIKVRENTSPKANKEIIDEAIVMAGVQSP 774
Db 721 RKVKVLGSAGFTVYKGIWIPDGENVKIPVAIKVRENTSPKANKEIIDEAIVMAGVQSP 780
QY 775 YVSRLLGICLTSTVOLVTQLMPYGCCLLDHVRNRRGLSGQDLLNWCMIKAGMSYLEDVR 834
Db 781 YVSRLLGICLTSTVOLVTQLMPYGCCLLDHVRNRRGLSGQDLLNWCMIKAGMSYLEDVR 840
QY 835 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIKMALESILRRFT 894
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIKMALESILRRFT 900
QY 895 HQSDVMSVGVTVWELMTGAKPYDGI PAREIPDLLEKGERLPQPPICITIDVYMINVKWM 954
Db 901 HQSDVMSVGVTVWELMTGAKPYDGI PAREIPDLLEKGERLPQPPICITIDVYMINVKWM 960
QY 955 IDSECRPRFRELVSFESFARMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDGMDGLVDA 1014
Db 961 IDSECRPRFRELVSFESFARMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDGMDGLVDA 1020
QY 1015 BEYLVPQGFPCDPAPGAGMWHRRSSSTRSGGDLTLGLEPSEEAAPSLAPSEG 1074
Db 1021 BEYLVPQGFPCDPAPGAGMWHRRSSSTRSGGDLTLGLEPSEEAAPSLAPSEG 1080
QY 1075 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVAPLTCSPQPEV 1134
Db 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVAPLTCSPQPEV 1140
QY 1135 NQDVRPQPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAGGAVENPEYLTQ 1194

Db 1021 EYLVPOQGFCDPAPAGAGMWHRRSSSTRSGGDLTLGLEPSEEAAPSLAPSEG 1080
QY 1075 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSTGTVAPLTCSPQPEYV 1134
Db 1081 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSTGTVAPLTCSPQPEYV 1140
QY 1135 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVYKDVFAFGGAVENPEYLTPO 1194
Db 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERPKTILSPGKNGVYKDVFAFGGAVENPEYLTPO 1200
QY 1195 GGAAPQHPHPPAPSPAFDNLYYWDQDPPERGAPSTFKGTPTAENPEYLGIDVPV 1249
Db 1201 GGAAPQHPHPPAPSPAFDNLYYWDQDPPERGAPSTFKGTPTAENPEYLGIDVPV 1255

RESULT 11

AAE24067
ID AAE24067 standard; Protein; 1255 AA.
AC AAE24067;
XX
XX 23-SEP-2002 (first entry)
XX Human Her-2 protein.
DE
XX Human; Her-2; epidermal growth factor receptor 2; infection; cancer;
KW hyperproliferative disorder; prophylaxis; inflammation; antisense;
KW tumour; gene therapy; phosphorothioate backbone.
XX Homo sapiens.
OS
XX MO200222636-A1.
PN
XX 21-MAR-2002.
PD
XX 12-SEP-2001; 2001WO-US28572.
PF
XX 15-SEP-2000; 2000US-0663834.
PR
XX (ISIS-) ISIS PHARM INC.
PA
XX Bennett CF, Cowsett LM;
PI WPI; 2002-471192/50.
XX N-PSDB; AAD38904.
DR
XX Novel antisense oligonucleotide which modulates the expression of Human
PT Epidermal Growth factor receptor, Her2, is useful for treating tumors
PT inflammation or to prevent infection in humans -
XX
XX Example 13; Page 95-107; 116pp; English.

CC The invention relates to antisense compounds targetted to a nucleic
CC acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)
CC that specifically hybridises with and inhibits the expression of Her2.
CC Antisense compounds of the invention are used for treating diseases or
CC conditions associated with Her2 such as hyperproliferative disorders
CC e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,
CC neural or cardiac cancer. They are also useful prophylactically e.g.
CC to prevent or delay infection, inflammation and tumour formation. The
CC invention is also used in gene therapy. The present sequence is human
CC Her-2 protein.

XX Sequence 1255 AA;

Query Match 98.5%; Score 6672; DB 23; Length 1255;
Best Local Similarity 98.2%; Pred. No. 0;
Matches 1232; Conservative 5; Mismatches 12; Indels 6; Gaps 1;
QY 1 MELAALCRNGLLALLPPGAASQVCT-----QYIKANSKFIGITELLYQCCVQGNL 54
Db 1 MELAALCRNGLLALLPPGAASQVCTGTDMLKRLPASPEETHLMLRHLYQCGVQVQGNL 60

QY 55 ELTYLPTNASLSFLQIOEYQGVYLIHAHQVQVPLQRLRIVRGTQLFEDNYALAVLDNG 114
Db 61 ELTYLPTNASLSFLQIOEYQGVYLIHAHQVQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
QY 115 DPLNNTTPTVGASPGGLRELQRLSLEILKGGVLIQRPOLCYQDTILWKDIFHKKNOLA 174
Db 121 DPLNNTTPTVGASPGGLRELQRLSLEILKGGVLIQRPOLCYQDTILWKDIFHKKNOLA 180
QY 175 LTLIDTNSRACHPCSPCKSGRCWGESSEDCQSLTRTVACGGCARGKGPLTDCCHEQC 234
Db 181 LTLIDTNSRACHPCSPCKSGRCWGESSEDCQSLTRTVACGGCARGKGPLTDCCHEQC 240
QY 235 AAGCTGPKHSDCLACLFHFNHSGICBLHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 294
Db 241 AAGCTGPKHSDCLACLFHFNHSGICBLHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 295 YNYLSTDVGSCTLVCPHNLNQEVTAEDGTQRCBKSKPCARVCYGLGMEHLREVRVTSAN 354
Db 301 YNYLSTDVGSCTLVCPHNLNQEVTAEDGTQRCBKSKPCARVCYGLGMEHLREVRVTSAN 360
QY 355 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFETLEETGYLYISAMPDLSLP 414
Db 361 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFETLEETGYLYISAMPDLSLP 420
QY 415 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLRELGLALIHNTHLFCFVHTV 474
Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLRELGLALIHNTHLFCFVHTV 480
QY 475 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHOLCARGHCGPGTQCVCNCSQFLRGQSC 534
Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHOLCARGHCGPGTQCVCNCSQFLRGQSC 540
QY 535 VEECEVLOGLPREYUNARHCLPCHPECOPOGNSVTCFGEADQCVACAHYKDPFPCVARC 594
Db 541 VEECEVLOGLPREYUNARHCLPCHPECOPOGNSVTCFGEADQCVACAHYKDPFPCVARC 600
QY 595 PSGVKPDLSYMPIWKFPDEEGACQPCPINCSTHSCVDLDDKGPAPQORASPLTSISAVVG 654
Db 601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCSTHSCVDLDDKGPAPQORASPLTSISAVVG 660
QY 655 ILLVVVLGVFGILIKRQOKIRKYTMRLRLQETELVEPLTPSGAMPNQAQMRILKETEL 714
Db 661 ILLVVVLGVFGILIKRQOKIRKYTMRLRLQETELVEPLTPSGAMPNQAQMRILKETEL 720
QY 715 RKVKYLGSGAGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAVVMAGVQSP 774
Db 721 RKVKYLGSGAGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAVVMAGVQSP 780
QY 775 YVSRLLGLCLTSTVOLVTQLMPYGCLLDHVRENRLGSLQDILLNWCMIKAGMSYLEDVR 834
Db 781 YVSRLLGLCLTSTVOLVTQLMPYGCLLDHVRENRLGSLQDILLNWCMIKAGMSYLEDVR 840
QY 835 LVHRDLAARNVLKSPNHNKVTDFGLARLLDIDETEHADGKVPDKWMALESILRRRPT 894
Db 841 LVHRDLAARNVLKSPNHNKVTDFGLARLLDIDETEHADGKVPDKWMALESILRRRPT 900
QY 895 HQSDVMSYGVTVWELMTGAKPYDGIIPAREIPDLLEKGERLPQPPICITDVMYIMVKWM 954
Db 901 HQSDVMSYGVTVWELMTGAKPYDGIIPAREIPDLLEKGERLPQPPICITDVMYIMVKWM 960
QY 955 IDSECRPFRELVSFSSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMDGLVDA 1014
Db 961 IDSECRPFRELVSFSSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMDGLVDA 1020
QY 1015 BEYLVPQGGFFCPDPAPGAGMWHRRSSSTRSGGDLTLGLEPSEEAAPSLAPSEG 1074
Db 1021 BEYLVPQGGFFCPDPAPGAGMWHRRSSSTRSGGDLTLGLEPSEEAAPSLAPSEG 1080
QY 1075 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSTGTVAPLTCSPQPEYV 1134
Db 1081 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSTGTVAPLTCSPQPEYV 1140
QY 1135 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVYKDVFAFGGAVENPEYLTPO 1194

Db 1021 EEYLVPPQGFCDPAPAGAGGMVHHRRSSSTRSGGDLTLGLEPSEEBAPRSLAPSRG 1080
QY 1075 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDTVPLPSTDCGYVAPLTCSPQPEYV 1134
Db 1081 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDTVPLPSTDCGYVAPLTCSPQPEYV 1140
QY 1135 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTLPSPGKNGVVDVFAFGGAVENPEYLTPO 1194
Db 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTLPSPGKNGVVDVFAFGGAVENPEYLTPO 1200
QY 1195 GGAAPQHPHPPAFSPAFDNLVYWDQPPRGPAPPSFTFKGTPTAENPEYLGLDVVP 1249
Db 1201 GGAAPQHPHPPAFSPAFDNLVYWDQPPRGPAPPSFTFKGTPTAENPEYLGLDVVP 1255

RESULT 13

AAW51143
ID AAW51143 standard; Protein; 1255 AA.

XX AAW51143;

XX 17-JUN-2002 (first entry)

XX Human Her-2/neu oncogene-encoded p185 glycoprotein.

XX Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;

KW tyrosine kinase; receptor; c-erbB2; gene therapy.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Domain 1..653

FT /note= "extracellular domain"

FT /note= "676..1255"

FT /note= "intracellular domain"

FT /note= "930..1255"

FT /note= "phosphorylation domain"

XX WO200212341-A2.

XX 14-FEB-2002.

XX 03-AUG-2001; 2001WO-US24283.

XX 03-AUG-2000; 2000US-0632507.

XX (CORI-) CORIXA CORP.

PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX Cheever MA, Gheysen D;

XX WPI; 2002-241743/29.

DR N-PSDB; ABA92250.

XX Her-2/neu fusion protein for treating or preventing cancer by eliciting

PT or enhancing an immune response to the protein, has Her-2/neu

PT extracellular domain fused to Her-2/neu intracellular or

PT phosphorylation domain -

XX Claim 68; Fig 7; 141pp; English.

XX The present sequence is that of human Her-2/neu (p185 glycoprotein
CC or c-erbB2), an oncogenic self-protein and target for anti-cancer
CC vaccines. The Her-2/neu gene is amplified and p185 is overexpressed
CC in a variety of cancers, including breast, ovarian, colon, lung and
CC prostate cancer. Her-2/neu is a member of the tyrosine kinase
CC family of receptor-like glycoproteins. It comprises an extracellular
CC domain with homology to the epidermal growth factor receptor
CC (EGFR), a highly hydrophobic transmembrane domain and a C-terminal
CC intracellular domain that also shows homology to EGFR. Its
CC overexpression correlates with a poor prognosis in breast and
CC ovarian cancers. The invention provides Her-2/neu fusion
CC proteins, nucleic acids encoding them, viral vectors, and vaccines

CC comprising the fusion proteins or nucleic acid molecules. In
CC preferred fusion proteins, the extracellular domain of a Her-2/neu
CC protein is fused to a Her-2/neu intracellular domain or
CC phosphorylation domain (or its DeltaRD fragment). An immune
CC response to Her-2/neu protein is elicited or enhanced by
CC administering the fusion protein in the form of a vaccine, or by
CC transfecting cells of an animal ex vivo with a nucleic acid
CC encoding the fusion protein, and delivering the transfected cells
CC to the animal. The fusion proteins, nucleic acids, and isolated
CC specific T-cells are useful for inhibiting the development of a
CC cancer, especially breast, ovarian, colon, lung or prostate cancer
CC in a patient. T cells that specifically react with a Her-2/neu
CC fusion protein can be used to remove tumour cells from a sample in
CC order to inhibit the development of cancer in a patient.

XX Sequence 1255 AA;

Query Match 98.5%; Score 6672; DB 23; Length 1255;

Best Local Similarity 98.2%; Pred. No. 0;

Matches 1232; Conservative 5; Mismatches 12; Indels 6; Gaps 1;

QY 1 MELAALCRWGLLLALLPFGAASQTQVCT-----QYIKANSKFEGITELLYQCCQVQGNL 54

Db 1 MELAALCRWGLLLALLPFGAASQTQVCTGCTDMKRLPASPETHLDMRLHYQCCQVQGNL 60

QY 55 ELTYLPTNASLSFLQDIOEVQGYVLIAHNOVROVPLQRLIRVGTQLFEDNYVALVDNG 114

Db 61 ELTYLPTNASLSFLQDIOEVQGYVLIAHNOVROVPLQRLIRVGTQLFEDNYVALVDNG 120

QY 115 DPLNNTTPTVGTASPGSLRELQRLSITELKGSVLFIQNPQLCYQDTILWKDIFHKNNQLA 174

Db 121 DPLNNTTPTVGTASPGSLRELQRLSITELKGSVLFIQNPQLCYQDTILWKDIFHKNNQLA 180

QY 175 LTLIDNRSRACHPCSPMKSGRCSWGESSEDCQSLTRTVCGAGCARCKGPLTDCHEQC 234

Db 181 LTLIDNRSRACHPCSPMKSGRCSWGESSEDCQSLTRTVCGAGCARCKGPLTDCHEQC 240

QY 235 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRTYTFGASCVTACP 294

Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRTYTFGASCVTACP 300

QY 295 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRVTSAN 354

Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRVTSAN 360

QY 355 IOEFAGCKKIFGSLAFIPESFDGDPASNTAPLOQVTFLEITGYLIYSAMPDLSLP 414

Db 361 IOEFAGCKKIFGSLAFIPESFDGDPASNTAPLOQVTFLEITGYLIYSAMPDLSLP 420

QY 415 DLSVFQNLQVIRGRILHNGAYSLTLQGLGSIWGLRSURELGSGLALIHNNTHLCFVHTV 474

Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGSIWGLRSURELGSGLALIHNNTHLCFVHTV 480

QY 475 PWDQLFRNPQHALLTANRPEDECYVGEGLACHQLCARGHCWGPPTQCVCNCSQFLRGQBC 534

Db 481 PWDQLFRNPQHALLTANRPEDECYVGEGLACHQLCARGHCWGPPTQCVCNCSQFLRGQBC 540

QY 535 VEECRVLOGLPREYVNAHSHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 594

Db 541 VEECRVLOGLPREYVNAHSHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600

QY 595 PSGVKPDI-SYMPIWKFPDEEGACQPCPINCSTHSCVDLDDKGCAPQASPLTSIVSAVVG 654

Db 601 PSGVKPDI-SYMPIWKFPDEEGACQPCPINCSTHSCVDLDDKGCAPQASPLTSIVSAVVG 660

QY 655 ILLVVVLGVFGILIKRQKIRKYTMRLRLLQETELVEPLTPSGAMPNQAQMRILKETEL 714

Db 661 ILLVVVLGVFGILIKRQKIRKYTMRLRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720

QY 715 RKVKVLGSGAGTGYVKGWIIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVSP 774

Db 721 RKVKVLGSGAGTGYVKGWIIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVSP 780

QY 775 YVSRLLGICLTSTVQLVTLQMPYGCILLDHHVNRNRLGSGDLLNWCQIAGKMSYLEDVR 834
DB 781 YVSRLLGICLTSTVQLVTLQMPYGCILLDHHVNRNRLGSGDLLNWCQIAGKMSYLEDVR 840
QY 835 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKKNWMALESILRRRT 894
DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKKNWMALESILRRRT 900
QY 895 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMVKWM 954
DB 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMVKWM 960
QY 955 IDSECPRELVSERMRDQRFVWIONEDLGPASPLDSTFYRSILDEDDMGDLVDA 1014
DB 961 IDSECPRELVSERMRDQRFVWIONEDLGPASPLDSTFYRSILDEDDMGDLVDA 1020
QY 1015 EYLVPOQGFCDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEERAPSPAPSEG 1074
DB 1021 EYLVPOQGFCDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEERAPSPAPSEG 1080
QY 1075 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1134
DB 1081 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
QY 1135 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVWVDVFAFGAVENPEYLTQ 1194
DB 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVWVDVFAFGAVENPEYLTQ 1200
QY 1195 GGAAPOPHPPAPSPAFDNLVYWDQPPERGAPPSTFKGTPTAENPEYLGLDVVP 1249
DB 1201 GGAAPOPHPPAPSPAFDNLVYWDQPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 14

AAU77114
ID AAU77114 standard; Protein; 1255 AA.
XX
AC AAU77114;
XX
DT 05-JUN-2002 (first entry)
XX
DE Human Her-2/neu polypeptide.
XX
KW Human; Her-2/neu; cytostatic; haematological malignancy; CML;
KW acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;
KW chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;
KW Hodgkin's lymphoma; T cell therapy.
XX
OS Homo sapiens.
XX
PN WO200213847-A2.
XX
PD 21-FEB-2002.
XX
PF 13-AUG-2001; 2001WO-US25408.
XX
PR 14-AUG-2000; 2000US-0638280.
PR 28-SEP-2000; 2000US-0675904.
XX
PA (CORI-) CORIXA CORP.
XX
PI Gaiger A, Cheever MA, Hand-zimmermann S;
FI WPI; 2002-280741/32.
DR N-PSDB; ABK10730.
XX
XX Inhibiting haematological malignancy development by administering
PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
PT encoding the polypeptide, or antigen presenting cells expressing the
PT polypeptide -
XX
PS Disclosure; Page 71-74; 74pp; English.
XX

CC The invention relates to a method for inhibiting development of
CC haematological malignancy in a patient by administering a polypeptide
CC comprising an immunogenic portion of Her-2/neu or a polynucleotide
CC encoding the polypeptide. Antigen presenting cells that express the
CC protein can also be administered. The sequences are used for inhibiting
CC development of haematological malignancy such as acute myelogenous
CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
CC lymphoma. This sequence represents the human Her-2/neu polypeptide.
XX
SQ Sequence 1255 AA;

Query Match 98.5%; Score 6672; DB 23; Length 1255;
Best Local Similarity 98.2%; Pred. No. 0;
Matches 1232; Conservative 5; Mismatches 12; Indels 6; Gaps 1;

QY 1 MELAALCRWGLLLALLPPCAASTVCT-----QYIKANSKFIGITELLYQCCVQVQNL 54
DB 1 MELAALCRWGLLLALLPPCAASTVCTGTDKMLRLPASPEHLDMLRHLYQCCVQVQNL 60
QY 55 ELTYLPTNASLSFLQDIOEVQGYVLIAHNVQVPLQRLIRIVRGTLQFEDNYALAVLNG 114
DB 61 ELTYLPTNASLSFLQDIOEVQGYVLIAHNVQVPLQRLIRIVRGTLQFEDNYALAVLNG 120
QY 115 DPLNNTTPTVGTASPGGLRELQRLSITELKGGVLIQRPOLCYQDTILWKDIFHKNNOLA 174
DB 121 DPLNNTTPTVGTASPGGLRELQRLSITELKGGVLIQRPOLCYQDTILWKDIFHKNNOLA 180
QY 175 LTLIDTNSRACHPCSPCKSGRSGESSEDCOSLTRTVCGGCARCKGPTDCCHQEC 234
DB 181 LTLIDTNSRACHPCSPCKSGRSGESSEDCOSLTRTVCGGCARCKGPTDCCHQEC 240
QY 235 AAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 294
DB 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 295 YNLSLTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRAVTSAN 354
DB 301 YNLSLTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRAVTSAN 360
QY 355 TOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOVFEETLEITGVLYISAMPDSL 414
DB 361 TOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOVFEETLEITGVLYISAMPDSL 420
QY 415 DLSVFQNLQVIRGRILHNGAYSLTLQGLIGISWGLRSRLGSLGALIHNNHLCFVHTV 474
DB 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLIGISWGLRSRLGSLGALIHNNHLCFVHTV 480
QY 475 PWDQLFRNPHQALLHTANRPEDECYGEGLACHQLCARGHCWGPPTQCVCNCSQFLRGQBC 534
DB 481 PWDQLFRNPHQALLHTANRPEDECYGEGLACHQLCARGHCWGPPTQCVCNCSQFLRGQBC 540
QY 535 VEECHVLOGLPREYVNAHCLPCHPECPONGSVTCFGEADOCVACAHYKDPFCVARC 594
DB 541 VEECHVLOGLPREYVNAHCLPCHPECPONGSVTCFGEADOCVACAHYKDPFCVARC 600
QY 595 PSGVKPDLSPYPIWKFPDEEGACQPCPINCSTHSCVDLDDKGPAPQORASPLTSIVSAVVG 654
DB 601 PSGVKPDLSPYPIWKFPDEEGACQPCPINCSTHSCVDLDDKGPAPQORASPLTSIVSAVVG 660
QY 655 ILLVVVLGVW?GILIKRRQOKIRKYTMRLRLQETELVEPLTPSGAMPNQAQMRILKETEL 714
DB 661 ILLVVVLGVW?GILIKRRQOKIRKYTMRLRLQETELVEPLTPSGAMPNQAQMRILKETEL 720
QY 715 RKVKVLGSGAGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVSP 774
DB 721 RKVKVLGSGAGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVSP 780
QY 775 YVSRLLGICLTSTVQLVTLQMPYGCILLDHHVNRNRLGSGDLLNWCQIAGKMSYLEDVR 834
DB 781 YVSRLLGICLTSTVQLVTLQMPYGCILLDHHVNRNRLGSGDLLNWCQIAGKMSYLEDVR 840
QY 835 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKKNWMALESILRRRT 894

Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMWALESIILRRRT 900
 Qy 895 HOSDWMSYGVWVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPCTIDVYMWKWM 954
 Db 901 HOSDWMSYGVWVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPCTIDVYMWKWM 960
 Qy 955 IDSECRPRFRELVSFSSMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDMDGLVDA 1014
 Db 961 IDSECRPRFRELVSFSSMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDMDGLVDA 1020
 Qy 1015 EYLVPOQGFCCPDPAAGAGMWHHRSSSTRSGGDLTLGLEPSEEBAPRSLAPSEG 1074
 Db 1021 EYLVPOQGFCCPDPAAGAGMWHHRSSSTRSGGDLTLGLEPSEEBAPRSLAPSEG 1080
 Qy 1075 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1134
 Db 1081 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
 Qy 1135 NOPDVRPQPPSPREGPLPAARPAATLERAKTLSPGKNGVVDVFAFGGAVENPEYLTPO 1194
 Db 1141 NOPDVRPQPPSPREGPLPAARPAATLERAKTLSPGKNGVVDVFAFGGAVENPEYLTPO 1200
 Qy 1195 GGAAPQHPHPPAFSPAFDNLVYWDQPPRGAPPSTFKGTPTAENPEYLGDLVPV 1249
 Db 1201 GGAAPQHPHPPAFSPAFDNLVYWDQPPRGAPPSTFKGTPTAENPEYLGDLVPV 1255

RESULT 15

AAR39568

ID AAR39568 standard; Protein; 1433 AA.

XX

AC AAR39568;

XX

DT 07-FEB-1994 (first entry)

XX

DE Sequence of c-erbB-2 tumour antigen.

XX

KW Tumour antigen; c-erbB-2; glycoprotein.

XX

OS Homo sapiens.

XX

XX

XX

XX

XX

PF 05-FEB-1993; 93WO-US01055.

XX

PR 06-FEB-1992; 92US-0831967.

XX

PA (CETU) CETUS ONCOLOGY CORP.

XX

PA (CREA-) CREATIVE BIOMOLECULES INC.

XX

PI Houston LL, Huston JS, Oppermann H, Ring DB;

XX

XX WPI; 1993-272889/34.

XX

XX N-PSDB; AAQ46083.

XX

PT New single chain Fv polypeptide binding to C-erbB-2 tumour
 PT antigen - for imaging or treating breast or ovarian cancer etc.
 XX
 XX Disclosure; pages 48-54; 87pp; English.
 XX
 CC c-erbB-2 refers to a protein antigen expressed on the surface of
 CC tumour cells. such as breast and ovarian tumour cells, which is an
 CC approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric
 CC pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39568 represents
 CC the location of a stop codon in AAQ46083.
 XX
 XX Sequence 1433 AA;
 SQ

Query Match

Best Local Similarity 97.9%; Score 6629; DB 14; Length 1433;

Pred. No. 0;

Matches 1225; Conservative 7; Mismatches 17; Indels 6; Gaps 1;
 Qy 1 MELAALCRWGLLALLPCCAASTOVCT-----QVYKANSKEFIGITTELYLQCCOVQGNL 54
 Db 1 MELAALCRWGLLALLPCCAASTOVCTGTDMKLRUPASPETHLDMLRLHYQCVQVQGNL 60
 Qy 55 ELTYLPTNASLFLQDIOEVQYVLI AHNQVRQVPLQRLRIVRGTLQDPEDNYALAVLDNG 114
 Db 61 ELTYLPTNASLFLQDIOEVQYVLI AHNQVRQVPLQRLRIVRGTLQDPEDNYALAVLDNG 120
 Qy 115 DPLNNTTPTVGTASPGGLRELQRLSLEILKGGVLIQRPQLCYQDTILWKDIFHKNNQLA 174
 Db 121 DPLNNTTPTVGTASPGGLRELQRLSLEILKGGVLIQRPQLCYQDTILWKDIFHKNNQLA 180
 Qy 175 LTLIDTNSRACHPCSPMCKSGRCWGESSEDCOSLTRITVACGCCARCKGPLTDCCHQEC 234
 Db 181 LTLIDTNSRACHPCSPMCKSGRCWGESSEDCOSLTRITVACGCCARCKGPLTDCCHQEC 240
 Qy 235 AAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 294
 Db 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
 Qy 295 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRAVTSAN 354
 Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRAVTSAN 360
 Qy 355 IOEFAGCKKIFGSLAPFESPDGDPASNTAPLOPELOVFEETLEETITGLYIISANPDSL 414
 Db 361 IOEFAGCKKIFGSLAPFESPDGDPASNTAPLOPELOVFEETLEETITGLYIISANPDSL 420
 Qy 415 DLSVFQNLQVIRGRILHNGAYSLTLQGLIGISWGLRSRELGSGLALHNNHLCFVHTV 474
 Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLIGISWGLRSRELGSGLALHNNHLCFVHTV 480
 Qy 475 PWDQLFNRPHQALLHTANRPEDECYGEGLACHQLCARGHCWGPGTQCVCNSQFIRGQEC 534
 Db 481 PWDQLFNRPHQALLHTANRPEDECYGEGLACHQLCARGHCWGPGTQCVCNSQFIRGQEC 540
 Qy 535 VEECHVLOGLPREYNARHCLPCHPECOPOGNSVTCFGEADOCVACAHYKDPFCVARC 594
 Db 541 VEECHVLOGLPREYNARHCLPCHPECOPOGNSVTCFGEADOCVACAHYKDPFCVARC 600
 Qy 595 PSGVKPDLSPYMPKPFDEBACQCPINCTHSCVDLDDKGPAPQASPLTISVAVVG 654
 Db 601 PSGVKPDLSPYMPKPFDEBACQCPINCTHSCVDLDDKGPAPQASPLTISVAVVG 660
 Qy 655 ILLVVVLGVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNOAQRILKETEL 714
 Db 661 ILLVVVLGVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNOAQRILKETEL 720
 Qy 715 RKVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEIIDEAYVMAGVSP 774
 Db 721 RKVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEIIDEAYVMAGVSP 780
 Qy 775 YVSRLLGICLTSTVOLVTLQMLPYGCLLDHVRNRRGLSQDILLNWCMIKAGMSYLEDVR 834
 Db 781 YVSRLLGICLTSTVOLVTLQMLPYGCLLDHVRNRRGLSQDILLNWCMIKAGMSYLEDVR 840
 Qy 835 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMWALESIILRRRT 894
 Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMWALESIILRRRT 900
 Qy 895 HQSDVMSYGVWVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPCTIDVYMWKWM 954
 Db 901 HQSDVMSYGVWVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPCTIDVYMWKWM 960
 Qy 955 IDSECRPRFRELVSFSSMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDMDGLVDA 1014
 Db 961 IDSECRPRFRELVSFSSMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDMDGLVDA 1020
 Qy 1015 EYLVPOQGFCCPDPAAGAGMWHHRSSSTRSGGDLTLGLEPSEEBAPRSLAPSEG 1074
 Db 1021 EYLVPOQGFCCPDPAAGAGMWHHRSSSTRSGGDLTLGLEPSEEBAPRSLAPSEG 1080

QY	1075	AGSDVFDGLGMAAKGLQSLPTHDPSPLOQYSEDTVPLPSETDGYVAPLTCSPQPEYV	1134
Db	1081	AGSDVFDGLGMAAKGLQSLPTHDPSPLOQYSEDTVPLPSETDGYVAPLTCSPQPEYV	1140
QY	1135	NOPDVRPQPPSPREGPLPAARPAATLERAKTILSPGKNGVVKDVFAFGGAVENPEYLTPO	1194
Db	1141	NOPDVRPQPPSPREGPLPAARPAATLERPKTILSPGKNGVVKDVFAFGGAVENPEYLTPO	1200
QY	1195	GGAAPQPHPPPAFSPAFDNLYYWDQPPPERGAPPSTFKGTPTAENPEYLGLDVPV	1249
Db	1201	GGAAPQPHPPPAFSPAFDNLYYWDQPPPERGAPPSTFKGTPTAENPEYLGLDVPV	1255

Search completed: July 22, 2003, 08:40:06
Job time : 42.7958 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:25:54 ; Search time 22.9062 Seconds
(without alignments)
5267.077 Million cell updates/sec

Title: SEQ4-59-73-12
Perfect score: 6809
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVPV 1255

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 73.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6729	98.8	1255	1 A24571	protein-tyrosine k
2	5919	86.9	1260	1 TVRTNU	protein-tyrosine k
3	5918.5	86.9	1254	2 I48161	p-185 precursor -
4	3117	45.8	1210	1 GQHUE	epidermal growth f
5	3093	45.4	1210	2 A53183	epidermal growth f
6	3085.5	45.3	1223	1 TVCHLV	epidermal growth f
7	2950.5	43.3	1308	2 A47253	epidermal growth f
8	2653	39.0	1166	1 S06142	protein-tyrosine k
9	2389.5	35.1	1342	2 A36223	kinase-related tra
10	2316.5	34.0	1339	2 JC4387	epidermal growth f
11	1786.5	25.9	698	1 TVFVLV	protein-tyrosine k
12	1703	25.0	604	1 TVYUHV	protein-tyrosine k
13	1654.5	24.3	1330	1 GQFFE	epidermal growth f
14	1647	24.2	544	2 S35745	protein-tyrosine k
15	1640	24.1	545	2 S00727	kinase-related tra
16	1623	23.8	540	2 B47776	protein-tyrosine k
17	1621	23.8	540	1 TVFVVB	protein-tyrosine k
18	1485	21.8	644	2 A36325	epidermal growth f
19	1275	18.7	1323	2 E88257	protein let-23 (im
20	1275	18.7	1374	2 S70712	protein-tyrosine k
21	1187	17.4	1369	2 S70713	protein-tyrosine k
22	1146	16.8	1717	1 A45558	epidermal growth f
23	1117	16.4	527	2 A42032	epidermal growth f
24	955.5	14.0	843	2 A27131	epidermal growth f
25	806.5	11.8	346	2 S13807	protein-tyrosine k
26	754.5	11.1	311	2 S13808	protein-tyrosine k
27	742	10.9	1363	2 T43220	insulin-like growt
28	717	10.5	1382	1 INHUR	insulin receptor p
29	710	10.4	1383	2 A36080	insulin receptor p

RESULT 1

A24571

Protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human

N/Alternate names: c-erb-B-2 protein precursor; kinase-related transforming protein erbB;
C/Species: Homo sapiens (man)

C/Date: 25-Oct-1987 #sequence.revision 06-Dec-1996 #text change 11-Jun-1999

C/Accession: A24571; A25491; A44188; B44188; I59509; I57622

R/Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T.; ; Nature 319, 230-234, 1986

A/Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth f.

A/Reference number: A24571; MUID:86118663; PMID:3003577

A/Accession: A24571

A/Molecule type: mRNA

A/Residues: 1-1255 <YAM>

A/Cross-references: GB:X03363; NID:G31197; PIDN:CAA27060.1; PID:G31198

R/Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.

Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985

A/Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epidermal

A/Reference number: A25491; MUID:86016729; PMID:2995967

A/Accession: A25491

A/Molecule type: DNA

A/Residues: 737-1031 <SEM>

A/Cross-references: GB:M11767; NID:G182163; PIDN:AAA35808.1; PID:G553282

R/Coussens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg, P.

Science 230, 1132-1139, 1985

A/Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chromos

A/Reference number: A44188; MUID:86070181; PMID:2999974

A/Accession: A44188

A/Molecule type: DNA

A/Residues: 740-910 <COU>

A/Cross-references: GB:M12036; NID:G183988; PIDN:AAA35978.1; PID:G183989

A/Accession: B44188

A/Molecule type: mRNA

A/Residues: 1-517, 'RALL', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>

A/Cross-references: GB:M11730; NID:G183986

R/King, C.R.; Kraus, M.H.; Aaronson, S.A.

Science 229, 974-976, 1985

A/Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.

A/Reference number: I59509; MUID:85272597; PMID:2992089

A/Accession: I59509

A/Status: translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 832-909 <REX>

A/Cross-references: GB:L29395; NID:G459807; PIDN:AAA35809.1; PID:G459808

R/Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.

Mol. Cell. Biol. 7, 2597-2601, 1987

A/Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptional

A/Reference number: I57622; MUID:87286898; PMID:3039351

A/Accession: I57622

A/Status: translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-191 <TAL>

A;Cross-references: GB:M16792; NID:q183983; PIDN:AAA58637.1; PID:g553332
C;Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
C;Genetics:
A;Gene: GDB:ERBB2; NGL; NEU; HER-2
A;Map position: 17q21.1-17q21.1
A;Introns: 25/1; 75/3; 147/1; 883/3
A;Note: the list of introns is incomplete
C;Function:
A;Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
inase
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>
F:722-653/Domain: extracellular #status predicted <EXR>
F:770-304/Domain: EGF receptor extracellular domain repeat <EE1>
F:395-605/Domain: EGF receptor extracellular domain repeat <EE2>
F:654-675/Domain: transmembrane #status predicted <TM>
F:676-1255/Domain: intracellular #status predicted <INT>
F:718-983/Domain: protein kinase homology <KIN>
F:728-734/Region: protein kinase ATP-binding motif
F:168,124,187,259,530,571,629/Binding site: carboxydrate (Asn) (covalent) #status predicted
F:686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:753/Active site: Lys #status predicted
F:1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 98.8%; Score 6729; DB 1; Length 1255;
Best Local Similarity 98.8%; Pred. No. 1.6e-268;
Matches 1240; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLLALLPPGAASCTCTDMMKRLPASPETHDMLRLHYOGCQVQGNL 60
Db 1 MELAALCRWGLLLALLPPGAASCTCTDMMKRLPASPETHDMLRLHYOGCQVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIOEVQOYIKANSKFIGITELQRLRIVRGTLQFPEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIOEVQOYIAHNOVQVPLRQLRIVRGTLQFPEDNYALAVLDNG 120
Qy 121 DPLNTTPTVGTASPGGLRELQRLSUTETILKGGVLIQRNPOLCYQDTILWKDIFHNQOLA 180
Db 121 DPLNTTPTVGTASPGGLRELQRLSUTETILKGGVLIQRNPOLCYQDTILWKDIFHNQOLA 180
Qy 181 LTLDTNRSRACHPCSPKCSRGWSESDCOSLTRVTCAGGACRCKGLPTDCCHQC 240
Db 181 LTLDTNRSRACHPCSPKCSRGWSESDCOSLTRVTCAGGACRCKGLPTDCCHQC 240
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 300
Qy 301 YNYLSTVGSCTLVCLNHNQVETADGTQRCCKSPCARVCYGLGMHRLREVRAVTSAN 360
Db 301 YNYLSTVGSCTLVCLNHNQVETADGTQRCCKSPCARVCYGLGMHRLREVRAVTSAN 360
Qy 361 IOEFAGCKKIFGSLAFLPESDGPASNTAPLOEQLOVFEITELCYLISAWPDSLP 420
Db 361 IOEFAGCKKIFGSLAFLPESDGPASNTAPLOEQLOVFEITELCYLISAWPDSLP 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSLTQGLGISWGLRSLRELGSGLALHNNHILCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSLTQGLGISWGLRSLRELGSGLALHNNHILCFVHTV 480
Qy 481 PWDQFRNPHOALLHTANRPEDCEVGEGLACHQLCARGHCWGPQTQCVCNCSQFLRGEC 540
Db 481 PWDQFRNPHOALLHTANRPEDCEVGEGLACHQLCARGHCWGPQTQCVCNCSQFLRGEC 540
Qy 541 VEECKVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFCVARC 600
Db 541 VEECKVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFCVARC 600
Qy 601 PSQVKPDLSPYMPKPFDEEGACQPCPINCTHSCVDLDDKGCQPAEQRASPLTISAVVG 660
Db 601 PSQVKPDLSPYMPKPFDEEGACQPCPINCTHSCVDLDDKGCQPAEQRASPLTISAVVG 660

Db 601 PSQVKPDLSPYMPKPFDEEGACQPCPINCTHSCVDLDDKGCQPAEQRASPLTISAVVG 660
Qy 661 ILLVVVLGVVFGILIKRQOKIRKYTMRRLQETELVEPLTPSGAMPNQAMRILKETEL 720
Db 661 ILLVVVLGVVFGILIKRQOKIRKYTMRRLQETELVEPLTPSGAMPNQAMRILKETEL 720
Qy 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKELDEAYVWAGVSP 780
Db 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKELDEAYVWAGVSP 780
Qy 781 YVSRLLIGLCTSTVQLVQVLMYPYGLLDHVRNRLGSGODLLNWCQIAKMSYLEDVR 840
Db 781 YVSRLLIGLCTSTVQLVQVLMYPYGLLDHVRNRLGSGODLLNWCQIAKMSYLEDVR 840
Qy 841 LVHRDLAARNVLYKSPNHVKITDPEGLARLDDIDETEHADGGKVPKIMMALESILRRFT 900
Db 841 LVHRDLAARNVLYKSPNHVKITDPEGLARLDDIDETEHADGGKVPKIMMALESILRRFT 900
Qy 901 HQSDVMSYGYVTWELMTFGAKPYDGPAREIPDLEKGERLPPOPICTIDVYIMVAKCM 960
Db 901 HQSDVMSYGYVTWELMTFGAKPYDGPAREIPDLEKGERLPPOPICTIDVYIMVAKCM 960
Qy 961 IDSECRPRFRELVSFSEFARMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDGMDLVA 1020
Db 961 IDSECRPRFRELVSFSEFARMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDGMDLVA 1020
Qy 1021 EYVLVPOOGFCDDPAPGAGMVHHRSSSTRSGGDLTLGLEPSEEEAPRSPPLASEG 1080
Db 1021 EYVLVPOOGFCDDPAPGAGMVHHRSSSTRSGGDLTLGLEPSEEEAPRSPPLASEG 1080
Qy 1081 AGSDVDFGDLGMAAKQLSLPTHDPSPLOYSBPTVPLPSETDGVVAPLTCSPQPEYV 1140
Db 1081 AGSDVDFGDLGMAAKQLSLPTHDPSPLOYSBPTVPLPSETDGVVAPLTCSPQPEYV 1140
Qy 1141 NQDVRPQPSPREGLPAAAPAGATLERAKTILSPGKGVVQKVFAGFGAVENPEYLTPO 1200
Db 1141 NQDVRPQPSPREGLPAAAPAGATLERAKTILSPGKGVVQKVFAGFGAVENPEYLTPO 1200
Qy 1201 GGAAPQHPHAPSPAFDNLVYWDODPPERGAPSTFKGTPTAENPEYGLDVPV 1255
Db 1201 GGAAPQHPHAPSPAFDNLVYWDODPPERGAPSTFKGTPTAENPEYGLDVPV 1255

RESULT 2
TVRNU
Protein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 11-Jun-1999
C;Accession: A24562; A61204
R;Bargmann, C.I.; Hung, M.C.; Weinberg, R.A.
Nature 319, 226-230, 1986
A;Title: The neu oncogene encodes an epidermal growth factor receptor-related protein.
A;Reference number: A24562; MUID:86118662; PMID:3945311
A;Accession: A24562
A;Molecule type: mRNA
A;Residues: 1-1260 <BAR>
A;Cross-references: EMBL:X03362; NID:g56745; PIDN:CAA27059.1; PID:g56746
R;Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohen,
Carcinogenesis 12, 1975-1978, 1991
A;Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals no m
2-thiazolyl]formamide or N-methyl-N-nitrosourea.
A;Reference number: A61204; MUID:92035293; PMID:1682063
A;Accession: A61204
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 637-663,'V',665-702 <MAS>
A;Note: authors translated the codon GCA for residue 25 as Val
C;Genetics:
A;Gene: neu
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>

A:Molecule type: mRNA
 A:Residues: 969-971, 'K', 973-1115, 'D' <EIS>
 A:Cross-references: EMBL:Z12608
 R:Heiseremann, G.J.; Gill, G.N.
 J. Biol. Chem. 263, 13152-13158, 1988
 A:Title: Epidermal growth factor receptor threonine and serine residues phosphorylated
 A:Reference number: A28941; MUID:88330814; PMID:3138233
 A:Accession: A28941
 A:Molecule type: protein
 A:Residues: 689-694, 'X', 696-704, 'L', 706-707, 989-992, 'XX', 995-996, 'X', 998-1000, 1002-1009,
 R:Hibbs, M.L.; Dunn, A.R.; Alexander, W.S.
 submitted to the EMBL Data Library, April 1994
 A:Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor
 A:Reference number: S45325
 A:Accession: S45325
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-971, 'K', 973-1210 <VER>
 A:Cross-references: EMBL:X78987; NID:G488830; PIDN:CAA55587.1; PID:G488831
 R:Paria, B.C.; Das, S.K.; Andrews, G.K.; Dev, S.K.
 Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993
 A:Title: Expression of the epidermal growth factor receptor gene is regulated in mouse b
 A:Reference number: 149643; MUID:93126380; PMID:7678348
 A:Accession: 149643
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 12-20, 22-132 <RES>
 A:Cross-references: GB:L06864; NID:G193001; PIDN:AAA53029.1; PID:G567201
 C:Genetics:
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphop
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:648-670/Domain: transmembrane #status predicted <TM>
 F:712-977/Domain: protein kinase homology <KIN>
 F:720-728/Region: protein kinase ATP-binding motif
 F:680, 695/Binding site: phosphate (Thr) (covalent) #status experimental
 F:993/Binding site: phosphate (Ser) (covalent) #status experimental
 F:993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental
 F:1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental
 F:1137/Binding site: phosphate (Tyr) (covalent) #status experimental

Query Match 45.4%; Score 3093; DB 2; Length 1210;
 Best Local Similarity 49.1%; Pred No. 1, 2e-119;
 Matches 625; Conservative 170; Mismatches 367; Indels 110; Gaps 23;

Qy 11 LLLALLPGAA--SQVCTGTDMLKRLPASPTHLDMLRHLYQCGVQVQNGLEITYLPTN 68
 Db 14 LTLALCAAGGALEKKVCQGTNSRLTQLGTPEHFLSLQRYNNCEVVLGNLEITYVQRN 73
 Qy 69 ASLSFLQDIOEVOQYIKANSFEIGTELQRLRIVRGTOIFEDNYALAVLDNGDPLNNTTP 128
 Db 74 YDLSEFKTQIEVAGYVLLALNTVERIPLENLQIRGNALYENTYALATLSN----- 124
 Qy 129 VTGASPGGLRELQRLSLTEILKGGVLIQKNPOLCYQDITLWKDI----FHKNNOLALTLI 184
 Db 125 -YGNRTGLRELPMNLQELLIGAVRFSNNPILCNMTDIQRDIVQNVFMSNMSDL---- 180
 Qy 185 DTRNSRACHPCSPMCKGRWGSESSDCQSLTRTVACGCA-RCKGPLPTCCHEQCAAG 243
 Db 181 -QSPSSCPKCDPSCPNCGSCWGGGBCNQKLTIIQAQCSHRCGRSPSCCHNQCAAG 239
 Qy 244 CTGPKHSDCLAHNHSIGICELHCPALVTYNTDTFESMPNPEGYTFGASCVTACPNY 303
 Db 240 CTGPRESCLVCQKQFDEATKDOTCPPLMLYNPTTYQMDVNPPEGYSFGATCVKCKPNY 299
 Qy 304 LSTDVGSCTLVCLPHNQEVTAEDGTQRCCKSKPCARCYGLGMEHLREAVTASNTQIE 363
 Db 300 VYTHGSCVRACGPPYIEV-EDGIRKCKKCDGPKRVNGIGIGEPKDTLSINATNKH 358
 Qy 364 FAGCKKIFGSLAFLPESFDGPASNTAPLOPEQLQVFTLEITGYLVISAWPDSLPDLS 423
 Db 359.FKYCTAISGDLHLPLVAFKGSFTPTPLDPRELEILKTVKEITGFLLIQAWPNWTDLH 418

RESULT 6

TVCHLV

epidermal growth factor receptor precursor - chicken

N;Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB

C;Species: Gallus gallus (chicken)

C;Date: 28-Feb-1996 #sequence_revision 05-May-1995 #text_change 04-Feb-2000

C;Accession: A27720; A00643

R;Lax, I.; Johnson, A.; Howk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Vennstr

Mol. Cell. Biol. 8, 1970-1978, 1988

A;Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mou

A;Reference number: A27720; MUID:88261272; PMID:3260329

A:Accession: A27720
A:Molecule type: mRNA
A:Residues: 1-1223 <LAX>
A:Cross-references: GB:M20386
R:Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1985
A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A:Reference number: A00643; MUID:85228222; PMID:2988784
A:Accession: A00643
A:Molecule type: mRNA
A:Residues: 585-1223 <NIL>
A:Cross-references: GB:M10066
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
C:Specific protein kinase
F:1-30/Domain: signal sequence #status predicted <SIG>
F:31-1223/Product: epidermal growth factor receptor #status predicted <MAT>
F:31-654/Domain: extracellular #status predicted <EXT>
F:81-307/Domain: EGF receptor extracellular domain repeat <EE1>
F:397-610/Domain: EGF receptor extracellular domain repeat <EE2>
F:655-677/Domain: transmembrane #status predicted <TM>
F:678-1223/Domain: intracellular #status predicted <INT>
F:719-984/Domain: protein kinase homology <KIN>
F:727-735/Region: protein kinase ATP-binding motif
F:136,202,280,361,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent) #s
F:192,650/Binding site: carbohydrate (Ser) (covalent) #status predicted
F:687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:754/Active site: Lys #status predicted
F:1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat

Query Match 45.3%; Score 3085.5; DB 1; Length 1223;
Best Local Similarity 48.1%; Pred. No. 2.5e-119;
Matches 624; Conservative 177; Mismatches 351; Indels 145; Gaps 25;

Qy 8 RWGLLLALLPGAA-----STQCTGDMKRLRLPASPETHDLMLRHLQCCVVOGNLE 61
Db 13 RGAALVLLLLGLVALCSAVEKVKVCGQTNNKLTQLGHVEDHFTSLQRMNNCEVVLNLE 72
Qy 62 LTYLPTNASLFLQDIOEQVOYIKANSFGITELQRLRIVRGTLQEDNEDYALVLDNGD 121
Db 73 ITYVHRNRLTFLKTOIEVAGYVLIAMNMVDVPLENLQIIRGNVLVDSFALAVLSNYH 132
Qy 122 PLNNTTPTVGTASPGGLRLQLRLSLEILKGGVLVORNPOLCYQDTILWKDIFHKNNOLAL 181
Db 133 -MNKTQ-----GLRELPMKRLSEILNGVVKLSNNPKLCNMDTVLNDIIDSRLK-PL 182
Qy 182 TLID -TNRSRACHPCSPMKSGRCWGSSESDCQLTRTVACGCA-RCKGPLPTDCCHEQ 239
Db 183 TVLDFASNLSSCPKCHPNCTEDHCWAGEQNCQTLTKVICAQCCSGRCGRKVPSDCCHQ 242
Qy 240 CAAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDTTFESMPNPEGRYTFGASCVTAC 299
Db 243 CAAGCTGPRSDCLACHFRDQATCKDTCPLVLYNPPTYQMDVNPPEGKYSFGATCYREC 302
Qy 300 PNYLSTDVSGCTLVCPILHNEVTAEDGTORCEKSPCARVCYGLGMEHLREVRVTS 359
Db 303 PHNYVTDHSGVRSVSCNTDYEV-EENGVRCKKCDGLCKVCMGIGELKGLISINAT 361
Qy 360 NIQFAGCKKIFGLAFLPESFDGDPASNTAPLOEQVETLEETIGLYISAWPDSL 419
Db 362 NIDSKNCTKINGDVSLPVAFLGDAFTKTLPLDPKLDVFTVKEISGFLLIQAWFDNA 421
Qy 420 PLSVFQNLQVIRGILHNGAYSITLQGLGISWLGSLRLSLGSLGLIHHNTHLCFVHT 479
Db 422 TDLYAFENLEIRTRKQHQYSLAVNLKIQSLGLSLSLKEISDGDIAIMKNKNCVADT 481
Qy 480 VPWDQLFNPHOALLHTANRDEDCVGEGLACHOLCARGHCWGPQTCVNCOSFLRQE 539
Db 482 MNWRSLFATOSQTKTIQNRRNNDCTADRHVCDPLCSDVGWCGPFPCHFCRSFSPKE 541
Qy 540 CVEECRVQLQLPREYVVARHCLPCHPECPQNG---SVTCFGRPADQCVACAHKDPFFC 596

Db 542 CVKQCNILQGEPRFERDSKCLPCHSECLVQNSTAYNTTCSGPGPDHCKMCAHFIDGPHC 601
Qy 597 VARPCSGVKPDLSPYMPKLPDEEGACQPCPINCSTHSCVDLDDKGCQAEQASPLTSTVS 656
Db 602 VKACPAGVLGENDTL-VMKYADANAVCOLCHPCNCRGCKGPGLEGCP---NGSKTPTIAA 657
Qy 657 AVV-GILLVVVLGVVFGILIKRRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAQRIL 715
Db 658 GWVGLLCLVVGGLGIGLYLRR-HIVRKRTLRLQLERELVEPLTPSGEAPNQAHLRIL 716
Qy 716 KETELRKVKVLGSGAFCTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYMA 775
Db 717 KETEFKVKVVLGSGAFCTVYKGLMIPGEKVKIPVAIKELREATSPKANKEILDEAYMA 776
Qy 776 GVSGPYVRLILGLCTSTVOLVLTOLMPYGLLQHVHNRGRGLSGODLLNMCWIAKMSY 835
Db 777 SVDPNPHVRLILGLCTSTVOLITQLMYPYGLLDVIREHKONIGSYQLLNMCVQIAKGNV 836
Qy 836 LEDVRLVHRLAARNVLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKMWALLESIL 895
Db 837 LEERLVHRLAARNVLVKTPQHVKITDFGLAKLGLGADEKEYHAEGKVPKMWALLESIL 896
Qy 896 RRPETHQSDVWSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPPPICTTIDVYIM 955
Db 897 HRIYTHQSDVWSYGVTVWELMTFGSKPYDGPAREIPDLLEKGERLPPPICTTIDVYIM 956
Qy 956 VKCWMIDSECRPFRELVSFSRWARDPQSFVVIQ-NEDLGPASPLDSTFVRSILLEDDMM 1014
Db 957 VKCWMIDADSRPKRELIAFSKWARDPPRYLVYIQGDERMHLPSPTDSKFYRTLMBEDM 1016
Qy 1015 GDVDAEYLVPOQGFCDPAPGAGGMVHHRSSSTRSGGDLTLGLEPSEEAERSP 1074
Db 1017 EDIVDAEYLVPHQGF-----NSPST-----SRTP 1042
Qy 1075 L-----APSEAGSDVFDGDLGMCAKGLQSLTHDPSPLQRYSEDTPVLPSET--DG 1127
Db 1043 LLSLSLSTNNSATNCID-----RNGQHPVRESFVQRYSSDPTGNFLEESIDDF 1094
Qy 1128 VAPLTCSPQEVYNQDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVKVDF-- 1185
Db 1095 L-----PAPEYVQ--LMPKPS-----TAMVQIYNNISLT 1125
Qy 1186 -----AFGAVENPEYLTPOGGAAQPPHPPAFSPAFDNLXYWQ----- 1225
Db 1126 AISKLPMDSRYQNSHSTAVDNPYEL-----NTNQSPKLVTFESSPTWISQNHQIN 1177
Qy 1226 -DPPE-----RGAPPSTFKGTPTAENPEYLGLDVP 1254
Db 1178 LDNPDYQDPLNETKPNGLLKVPAENPEYLRVAAP 1214

RESULT 7

A47253
epidermal growth factor receptor, HER4 - human
C:Species: Homo sapiens (man)
C>Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C:Accession: A47253
R:Flouman, G.D.; Culouscou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.; Ne
Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993
A:Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epidermal
A:Reference number: A47253; MUID:93189574; PMID:838336
A:Accession: A47253
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-1308 <PLO>
A:Cross-references: GB:I07868; NID:G337359; PIDN:AAB59446.1; PID:G337360
A>Note: sequence extracted from NCBI backbone (NCBIP:126842)
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; growth factor receptor
F:716-981/Domain: protein kinase homology <KIN>
F:724-732/Region: protein kinase ATP-binding motif

Query Match	43.3%;	Score	2950.5;	DB 2;	Length	1308;	
Best Local Similarity	45.0%;	Pred. No.	8.9e-114;	Indels	173;	Gaps	28;
Matches	606;	Conservative	183;	Mismatches	384;		
QY	9	WGLLLALLPPGAA-----STQVCTGTDMKRLRPAS	PTHLDMLRLHLYQGCQVVOGNLELTY	64			
DB	8	WVWSLLVAAGTVQPSDSQSVCAGTENKLSLSLEQOYRALRKYYENCEVVMGNLEITS	67				
QY	65	LPTNASLSFLQIDIEVOOYIKANSKFITIBLQRLRIVRGTFQLFEDNYALAVLDNGDPLN	124				
DB	68	IEHRDLISFLASVREVTGYVLVALNQRYFLPLENLRIIRGTKLIEDRYLAIFLNRYKDG	127				
QY	125	NTTPVTGASPGGLREQLRSITELKGGVLQORNPOLCYQDTILWKDIFHKNNQLALTLI	184				
DB	128	NF-----GLOELGKUNITEILNGVYVDQNKFLCYADTIHWQDIVRNWPNPSNLTLV	178				
QY	185	DTRNSRACHPCSPMKSKRCWGESSEDCQSILTRTVACGGC-ARCKGPIPTDCHEQCAAG	243				
DB	179	STNGSSGCRCHKSTG-RWCPTENHCQTLTRIVCAEQCDGRCVGPVSDCHRCAGG	237				
QY	244	CTGPKHSCLACLHFNHSGICELHCPALVTVNTDTFESMPNDEGRYTFGASCVTACPNY	303				
DB	238	CSGPKDTCFACMNFNDSGACVTCQPTQFVYNPTTFQLEHFNFAKTYTGAFCKVKCPHNF	297				
QY	304	LSTDVGSCTLVCLPHNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVAVTSANIQE	363				
DB	298	V-VDSSSCVRAPSPSKMEV-EENGIMCKPCTDIPCACDGIPTGSLMSAOTVDSNIDK	355				
QY	364	FAGCKIFGSLAFIPESPDGPASNTAPLOEQLOQVFTLEITGILYISAWPDSLPDLS	423				
DB	356	FINTCKINGNLIFLVTHGHPDYNIAEIAIDPEKLVNFTVREITGFLNIQSWPNNMTDFS	415				
QY	424	VFQNLQVIRGLIHNAGVSLTQGLIGISWLGRLSRLGSLALIHNNTHLCFVHTVPWD	483				
DB	416	VFSNLVTIGRVLVSGLSLLILKQGITSLQFSLSKISAGNIYITDINSICYHTTINWT	475				
QY	484	QLFRNPHQALLHTANRPEDECVGEGLACHOLCARGCHWGPGPTQVCNCSQFLRGQECVEE	543				
DB	476	TLFSTINQRIVIRDNRAKENTAEAGWCVNHLCCSDGCGWPGPDQCLSCRRFSRGRICIES	535				
QY	544	CRVLQGLPREVYNARHCLFCHPECOF-QNGSVTCFGEADOCVCAHYKDPFFCVARCP	602				
DB	536	CNLYDGEFREFENGSIQVCEPDQCEKMEGDLTCHGPGDNTCKSHKFDGPNCKVCPD	595				
QY	603	GVPKPLSYMPIWPFDEEGACQPCPINCTHSCVDLDDKGC-----PAEQRASPL	651				
DB	596	GLOGANSF--IFKYADPDRECHPCPNCTQCGNGPTSHDCIYVPWTGHTSLPQHAR-TPL	652				
QY	652	TSIVSAVV-GILLVVLGVVFGILIKRQOKIRKYTMERLLQETELVEPLTPSGAMPNOA	710				
DB	653	--IAAGVIGLFLVIVGLTFVAVYVRKSIK-KKRALRRFL-ETELVEPLTPSGTAPNOA	708				
QY	711	QMRILKTELKRVKVLGSAFTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE	770				
DB	709	QLRILKTELKRVKVLGSAFTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE	768				
QY	771	AYWAGVSPYVSRLLGICLITSTVOLVTQLMPYGLLDHVRNRLGSLQDLNWCQIA	830				
DB	769	ALIMASMDHPHLVRLGVCLSPITQLVTLQMPHGCLLEYVHEHKNIGNSQLLLNWCQIA	828				
QY	831	KGMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEXHADGSKVPKQMA	890				
DB	829	KGMVLEERLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEXHADGSKVPKQMA	888				
QY	891	LESILRRFTTHQSDVMSYGVTVWELMTFCAPYDGPAREIPDLLEKGERLPQPPICITD	950				
DB	889	LECIVHFKTHQSDVMSYGVTVWELMTFGKPYDGPAREIPDLLEKGERLPQPPICITD	948				
QY	951	VYIMVWKCMIDSECRPRELUSFESRMARDPQRFVQIONED-LGPASPLDSTFYRSL	1009				
DB	949	VYIMVWKCMIDSECRPRELUSFESRMARDPQRFVQIONED-LGPASPLDSTFYRSL	1008				
QY	1010	*EDDDMGDLVDAEYLVPOQGFPCPDAPCAGGVMVHRRSSSTRSGGDLTLGLPSEEE	1069				

Db	1009	DEEDLEDMDAEBYLP-QAFNIPPP-----IYTSRARIDSNRS-----EIGHSPPPAY	1056				
Qy	1070	APRS-----PLAP-SEGAGSDVDFDCLGMGAACLOS	1100				
Db	1057	TPMSGNQFVYRDGFAAEQGVSVYPRAPTSTIPEAPVAQGATAEIFFDDSCCNGTLRKPVA	1116				
Qy	1101	LPTHDSPLQRYSEDPTVPLPS-----ETDGVAPLTCSPQPEYVQPDVAPQPPSPR	1153				
Db	1117	PHVQDSSTQRYSDPTVFAPERSPGELDEEGMTIMRDKPKQEYLNPNVE-----	1167				
Qy	1154	EGPLPAARPAAGATLERAKTILSPGKNGVQVDFAFGGAVENPEYLT PQGGAAPQHPPPA-	1212				
Db	1168	ENFPVSR-----KNGDLQ-----ALDNPEYHNASNG-----PKAE	1199				
Qy	1213	-----FSAFDNLYYWDQDPPPERGA--PPSTF	1237				
Db	1200	DEYVNEPLYLNTFANTLGKAEYLNKNNILMSPEKAKAFDNPDMYNNHSLPRSTLQHPDYL	1259				
Qy	1238	KGTPT-----AENPEYL	1249				
Db	1260	QYSTKYFYKONGRIPIVAENPEYL	1285				
RESULT 8							
S06142							
Protein-tyrosine kinase (EC 2.7.1.112) mrk-Y precursor - southern platyfish							
N;Alternate names: epidermal growth factor receptor homolog; kinase-related transforming							
C;Species: Xiphophorus maculatus (southern platyfish)							
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 18-Feb-2000							
C;Accession: S06142; S13809							
R;Wittbrodt, J.; Adam, D.; Mailitschek, B.; Maeueller, W.; Raulf, F.; Telling, A.; Roberts							
Nature 341, 415-421, 1989							
A;Title: Novel putative receptor tyrosine kinase encoded by the melanoma-inducing Tu loc							
A;Reference number: S06142; MUID:90015140; PMID:2797166							
A;Accession: S06142							
A;Molecule type: DNA							
A;Residues: 1-1166 <WIT>							
A;Cross-references: EMBL:X16891; NID:G65290; PIDN:CAA34770.1; PID:G65291							
R;Adam, D.; Maeueller, W.; Scharlt, M.							
Oncogene 6, 73-80, 1991							
A;Title: Transcriptional activation of the melanoma inducing Xmrk oncogene in Xiphophoru							
A;Reference number: S13807; MUID:91125882; PMID:1846957							
A;Accession: S13809							
A;Status: preliminary; translation not shown							
A;Molecule type: DNA							
A;Residues: 821-1025, 'N', 1027-1098, 'A', 1100-1166 <ADA>							
A;Cross-references: EMBL:X56319; NID:G65284; PIDN:CAA39763.1; PID:G65285							
C;Genetics:							
A;Gene: mrk							
A;Map position: Y							
A;Introns: 872/3; 898/1; 947/1; 979/3; 1025/3; 1056/1							
C;Superfamily: epidermal growth factor receptor; protein kinase homology							
C;Keywords: ATP; growth factor receptor; phosphotransferase; transmembrane protein; tyro							
F;1-25/Domain: signal sequence #status predicted <SIG>							
F;26-1166/Product: kinase-related transforming protein (Tu) #status predicted <MAT>							
F;707-972/Domain: protein kinase homology <KIN>							
F;715-723/Region: protein kinase ATP-binding motif							
Query Match							
Best Local Similarity 44.8%; Pred. No. 1.2e-101;							
Matches 568; Conservative 163; Mismatches 398; Indels 138; Gaps 28;							
Qy	4	AALCRWGLLALLPPGAAT-----QVCTGTMKRLRPLASPETHLMDRLHLYQGCQVVOGN	59				
Db	8	AALLQ--LLLVLSIRCCSTDPDRKVCQGTSNQMT---LDNHYLKKMKMYSGCNVLEEN	62				
Qy	60	LEITYLPTNASLSFLQIDIEVOOYIKANSKFITIBLQRLRIVRGTFQLPEDNYVALAVLDN	119				
Db	63	LEITYTOENQDLSFLQISIEVGGVFLIAMEVSTIPLNLRKQNLQYEGNFTLLVWSN	122				
Qy	120	GDPLNNTTPTVTGASPGGLREQLRSITELKGGVLQORNPOLCYQDTILWKDIFHKNNQL	179				

Db 123 YQK-NPSSP--DVYQVGLKQLQLSNLTLISGGVKSHPNLLCNVETINWWDIVDKTSNP 179
Qy 180 ALTLIDNRSRACHPCSPMKSGRCWGESSEDCSLTRTVCAAGC-ARCKGPLPTDCHE 238
Db 180 TNNLIHAFERQCKQCDHGCYVNGSWAPGPGHCQKFTKLLCAEQCNRRCRGPKPIDCNE 239
Qy 239 QCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNBEGRYTFGASCVTA 298
Db 240 HCAGCGTGPATDCLACRDFNDGCTKDTCPPIKYDIVSHVQVVDNPNIKYTFGAACVKE 299
Qy 299 CPYNVLTGSGCTLVCPHMQEYTAEDGTQRCCKSPCARVCYGLGMEHLREVRVATS 358
Db 300 CPSNVVTE-GACVRSACAGMLEVD-ENGRKSCPKPCDVCVCKVCDGIGISLNTIAVNS 357
Qy 359 ANIQEFAGCKKIFGSLAFPLPSPDGPASNTAPLQPEOLQVFEETLEBITGYLISAMPDS 418
Db 358 TNRFSNCTKINGDIILNRNSFEGDPHYKIGTMDPEHLNLTIVKEITGYLVIMWPFEN 417
Qy 419 LPDLVSFQNLQVIRGRIHNGAYS-LTLQGLIGISWGLRSRLRELGLALHHTHLCFV 477
Db 418 MTSLSVFQNLBIIRGRTTFSFGFVVVQVVRHLQWGLRSLSKEYSAGNVILKNTLQRYA 477
Qy 478 HTVPWDOLFNRPHOALLHTANRPEDECVGEGGLACHOLCARGHCWGPPTQCVNCSQFLRG 537
Db 478 NTINWRILFRSEDSIEYDART-----ENQTCNNECSEDCGW-PGPTMCVSCUHVDRG 529
Qy 538 QECVEECRVLQGLPREYVNAHRLCHPECPQNGSVTCFGEADQCACAHYKDPFCV 597
Db 530 GRCVASCNLLQGEPREAQVDCRCVQCHQELVQDLSLTCYCGPANCKSAHFQDGPQCI 589
Qy 598 ARCPGKVPDLSYMPIWKPDEBACQPCPINCTHSCVDLDDKCPAEQASPLTSTVSA 657
Db 590 PRCPHGILGDDTL-IMKYADKMGCCQPCQNCQTCGSGPGLSGCRGD-IVSHSSLAVAL 647
Qy 658 VVGLLVVVLGVGILIKRQOKIRKYVTRILLQETLEVEPLTPSGAMPNQAOIRLKE 717
Db 648 VSGLIITVIVALLIVLRRRIK-RKRTIKLQELVEPLTPSGQAPNQAFRLIKE 706
Qy 718 TELRKVKLGSGAGFYVKGIIWIPDGENVKIPVAIKVLRNTPSKANKEIIDEAYVMAGV 777
Db 707 TEFKDRVLGSGAGFYVKGIIWIPDGENIRIPVAIKVLRNTPSKANKEIIDEAYVMAGV 766
Qy 778 GSPVSVLLGCLTSTVOLVTQMLPYGCLLDHVRNRRGLSGQDLLNWCQIAKGMYLE 837
Db 767 DHPHVCRLGCLTSAVLQVLTQMLPYGCLLDYVRQHERICQWLLNWCQIAKGMYLE 826
Qy 838 DVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIMWALESLRR 897
Db 827 ERHLVHRDLAARNVLLKNPNHVKITDFGLSKLLTADEKEYQADGGKVPKIMWALESLQW 886
Qy 898 RFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKERLPQPPICTIDVYIMWK 957
Db 887 TYTHQSDVMSYGVTVWELMTFGSKPYDGIIPAKETASVLENGERLPQPPICTIEVYIMLK 946
Qy 958 CWMIDSCRRFRFELVSEFSRMARDPQRFVVIQNEIDLGASPLDSTFYRSLLEDDMGDL 1017
Db 947 CWMIDPSRRFRFELVGEFSGMARDPSRYLVIQ--NLPSLSRRLFRSLSSDD--DV 1001
Qy 1018 VDAEYLVPOGGFCPPDAPGAGGVHHRSSSTRSGGGDLTLGLFSPSEEAAPRSLAP 1077
Db 1002 VDAEYLLPYKRI-----NRQGS-----EPCIP 1024
Qy 1078 SEGAGSDVFDGLGMAAGKGLQLPHTDPSPLQRYSEDPV-PLPSETDGVAPLTCSPQ 1136
Db 1025 PTGH-----PVRENSTILRNISDPTQNALEKLDGH----- 1055
Qy 1137 PEYVNPQDVRQP-----PSPRE-----GPLP-AARPAGATLAKTLSPGRNGYVKD 1183
Db 1056 -EYVNPQGETSSRLSDIYNPNVEDLTDGWPVSLSSQEAETNFSRPEYLNTNQSJ--- 1111
Qy 1184 VFAGGAVENPEYLTPOGGAAPQHPAPSPAFDNLYYWQDDPQPPGAPSPFKGTPTA 1243
Db 1112 PLVSSGSMDDPDY---QAG-----YQAAF-----LPOTGALTGNGMFLPAA 1149

Qy 1244 ENPEYLG 1250
Db 1150 ENLEYLG 1156

RESULT 9

A36223
kinase-related transforming protein (erBB3) (EC 2.7.1.-) precursor - human
C:Species: Homo sapiens (man)
C>Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000
C:Accession: A36223; I59164
R:Kraus, M.H.; Issing, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.
Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989
A:Title: Isolation and characterization of ERBB3, a third member of the ERBB/epidermal g
A:Reference number: A36223; MUID:90083234; PMID:2687875
A:Accession: A36223
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1342 <KRA>
R:Plowman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todaro, G.J
Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990
A:Title: Molecular cloning and expression of another epidermal growth factor receptor-re
A:Reference number: I59164; MUID:90311312; PMID:2164210
A:Accession: I59164
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-559, 'G', 561-957, 'F', 959-1063, 'G', 1065-1342 <RES>
A:Cross-references: GB:M29366
C:Genetics:
A:Gene: GDB:ERBB3; HER3
A:Cross-references: GDB:119880; OMIM:190151
A:Map position: 12q13-12q13
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo
C:Keywords: ATP; phosphotransferase
F:707-972/Domain: protein kinase homology <KIN>
F:715-723/Region: protein kinase ATP-binding motif

Query Match 35.1%; Score 2389, 5; DB 2; Length 1342;
Best Local Similarity 40.2%; Pred. No. 8.5e-91;
Matches 527; Conservative 189; Mismatches 466; Indels 129; Gaps 32;

Qy 10 GLLLALLPGAA--STQVCTGTDMLKRLPASPEHLRLHLYQGCGVOGQNLLETYLPT 67
Db 11 GLLFSLARGESEVNSQAVCPGTGLNSLVTGDAENQYQLYKLYERCEVVMGNLEIVLTGH 70
Qy 68 NASLSFLQDIQEVQYIKANSKFIGITELQRLRIVRGTLQFEDNYALAVLDNGDPLNNTT 127
Db 71 NADLSFLQWIREVTGYVLVAMNEFSTLPLNLRVVRGTQVYDGKFAIFVM----LYNNT 125
Qy 128 PVTGASPGGLRELQSLRLTEILKGGVLIQBNPOLCYQDITLWKDIPHKNNQLALTLDTN 187
Db 126 ----NSSHALRQLRLTQLTEILSGGVIYEKNDKLCMDTIDWRD---AEIVVKD 178
Qy 188 RSRACHPCSPMKSGRCWGESSEDCSLTRTVCAAGC-ARCKGPLPTDCHEQCAAGCTG 246
Db 179 NGRSCFPCEVCKG-RCWPGSEDCQTLTKTICAPQCNHGCFGNPNQCCHECAGCGSG 237
Qy 247 PKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNBEGRYTFGASCVTACPYNYLST 306
Db 238 PQDTCDFACRHFNDSGACVPRCPQLVYNKLTLEPSPHTKYQYGGVCVASCAPHNV-V 296
Qy 307 DVGSCCTLVCPHMQEYTAEDGTQRCCKSPCARVCYGLGMEHLREVRVATSANIOEPAG 366
Db 297 DOTSCVCRACPPDKMEVD-KNGLKMCPCGGLCPKACEGTGSG--SRFQTVDSNIDGFVN 353
Qy 367 CKKIFGSLAFPLPSPDGPASNTAPLQPEOLQVFEETLEBITGYLISAMPDSPLDSVFQ 426
Db 354 CTKILGNLDFLITGLNGDPWHKIPALDPEKLVNFRVREITGYLNTQSPHPHNFVSFS 413
Qy 427 NLOVIRGRILHNGAYS-LTLQGLIGISWGLRSRLRELGLALHHTHLCFVHTVPWDQL 485

Db 414 NLTTIGRSLYNRGSLIMKLNVTSLGRSLKEISAGRIYISANRQLCYHSLNWTKV 473
 Qy 486 FRNPHQALLHTA-NRPEDECVEGLACHOLCARGHCWGPQTQVNCVSOFLRGQECVEEC 544
 Db 474 LRGPTEERLDIKHNRPRDCAEAGKVCVDCPLSCSSGGCWGPGQCCLSCRNYSRGVGVCTHC 533
 Qy 545 RVLCGLPREYVYARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPDPCVRCPSGV 604
 Db 534 NFLNGEPREFAHEACSCHEPCQMEGTATCNGSDTCAQACHFRDGPCHVSCSPHGV 593
 Qy 605 KPDLSYMPIWKPDEEGACQPCINCHTHSCVDLDDKGCAPBQRA-----SPLTSTVSVAWG 660
 Db 594 LG--AKGPIYKYPDVQNECRPCHENCCTGCKGPELQDCLGQTLVIGKTHLTALTVIAG 651
 Qy 661 ILVAVLVGVFGILIKRQOKIR-KYTHRRLLQETELVEPLTPSGAMPNQAMQILKETE 719
 Db 652 --LVVIMFLMGTTFLYWRGRRIQNKRRMYRLERGESIEPLDPS--EKANKVLARIFKETE 708
 Qy 720 LRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMAGVGS 779
 Db 709 LRKVLGSGVFGTVHKGVWIPGESIKIPVCIKVIEDKSRQSFQAVTDHMLAIGSLDH 768
 Qy 780 PYVSRLLGICITSTVQLVTLQMPYGCCLLDHVRENRGLSGDOLLNWCMIQIAKMSYLEDV 839
 Db 769 AHIVRLGLCPGSSQLQVTLPLGLSLDHRQHGALGPOLLNWNQVQIAKGMVYLEEH 828
 Qy 840 RLVRDLAARNVLKSPNHVKITDFGLARILDDIETEHADGGKVPKIMMALESILRRFP 899
 Db 829 GMVHRNLAARNVLKSPSQVQVADFGVADLLPPDDKQLLYSEAKTPIKIMMALESIHFGKY 888
 Qy 900 THQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMVWKCV 959
 Db 889 THQSDVMSYGVTVWELMTFGAEPTAGLAEVPLDLEKGERLAQPOICTIDVYIMVWKCV 948
 Qy 960 MIDSECRPRFELVSEFSEFARMARDPQFVVIQNEDLGPA---SPLDSTFYRSLLEDDMDGD 1016
 Db 949 MIDENIRPTEKELANEFTRMARDPPRYLVIKRES--GPGIAPGPBPHGLTNKKLEVELEP 1007
 Qy 1017 LVDAEYVLVPOQGFPCPPAPGAGGVVHRRSSSTRSGGDLTLGLLEP--SEEAAPRSPL 1075
 Db 1008 ELDDLDDLEAED-----NLATTLGSLALPVGTLRNPRGSOLL 1048
 Qy 1076 APSEGAGSDVFDGLGMAAGLQSLPTH--PSPLQRYSEDPTVPLP-----SETDGYV 1128
 Db 1049 SPSSGY-WPMNQNLGSCQSAVSGSSERCPVSLH-----PMPRGCLASSESEGV 1101
 Qy 1129 A-----PLTCSPOPE-----YVNPQDVVRPQPPSPREGP-----L 1157
 Db 1102 TGSEAEQLQKVMCRSRSRSPRGDSAYHSQRHSLLTPTVPLSPGLEEDVNGYVM 1161
 Qy 1158 PAARPAGATLERAKTLSP-GKNGVV-----KDVPAFGAVENPEYLTPOGGAAPQPHPP 1210
 Db 1162 PDTHLKGTSPSSREGTLSSVGLSLVLTGEED-----EVEYMNRRRHSP-PHPP 1212
 Qy 1211 PAFSPAFDNLVYWD-----ODPPERGAPESTFKGTPTAENPEYL 1249
 Db 1213 RPSSLEELGEYMDVGSLSASLSTQSCPLHPVIMPTAGTTDDEYEM 1263

RESULT 10

JC4387
 epidermal growth factor receptor homolog precursor - rat
 N:Alternate names: ErbB3 protein; HER3 protein
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998
 C:Accession: JC4387
 R:Helleyer, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.
 Gene 165, 279-284, 1995
 A>Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protein.
 A:Reference number: JC4387; MUID:96096535; PMID:8522190
 A:Accession: JC4387
 A:Molecule type: mRNA
 A:Residues: 1-4339 <HEL>

A:Cross-references: GB:U29339; NID:g915389; PID:g915390
 A:Experimental source: liver
 A>Note: The authors translated the codon AAC for residue 369 as Thr and GTT for residue 370 as Thr. This protein is a functional heregulin receptor that transduces signals to the C:Genetics:
 A:Gene: ErbB3
 C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog; C:Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane protein F;1-19/Domain: signal sequence #status predicted <SIG>
 F;20-1339/Product: epidermal growth factor homolog #status predicted <WAT>
 F;640-659/Domain: transmembrane #status predicted <TMW>
 F;705-970/Domain: protein kinase homology <KIN>
 F;713-721/Region: protein kinase ATP-binding motif
 F;939,1051,1156,1194,1196,1219,1257,1259,1273,1286,1325/Binding site: phosphate (Tyr) (c

Query Match 34.0%; Score 2316.5; DB 2; Length 1339;

Best Local Similarity 40.4%; Pred. No. 8.3e-88;

Matches 518; Conservative 170; Mismatches 440; Indels 155; Gaps 34;

Qy 3 LAALCRWGLLALLPGAA---STQVCTGTDMLRLPASPTHLDMLRHLYQGCVVQGN 59
 Db 7 LQVLC-----FLLSLARGSEMNSQAVCPGLTGLSVTGDADNQYOTLYLYKECEVMGN 62
 Qy 60 LELTYLPTNASLFDIOEVQYIKANSKFEIGITELQRLIRVRGTOLFDENYALAVLDN 119
 Db 63 LEIVLTGHNADLSFLQWIREVYAVLVAMNEFSLPLNLRVVRGTQVYDGKFAIFVM-- 120
 Qy 120 GDPINNTPTVGASPGGLREQLRLSLTEILKGGVLIOQNPOLCYQDTILAKDIFHKNNQL 179
 Db 121 ---LNVT-----NSSHALRQLKFTQLTEILSGGYIEKDKLCHMDTIDRVIDRVR-- 170
 Qy 180 ALTLIDNRSRACHPCSPMCKSRGSESSDQSLTRTVCAAGC--ARCKGPLPTDCHE 238
 Db 171 GAELVKNNGANCPCHVECKG--RCWGPDPDQQLITKICAPQCNGRCGPNQCHD 229
 Qy 239 QCAAGCTGPHSDCLACLFHNSGICELHCPALVTYNTDTFESMNPENEGTYTGAASVTA 298
 Db 230 ECAGGCGSPQDTCFACRRFNDGACVPRCPPELVYKNTLFQLEPNPHTKYQGGVCVAS 289
 Qy 299 CPNYLSTDVGSCTVLCPLHNOVTAEDGTQRCCKSPCARVCYGLGMEHLREAVTS 358
 Db 290 CPHNFV--VDQTFVCRACFPDKMEVD--KHGLKMCPECGGLCPKACEGTGSG--SRYTQVDS 345
 Qy 359 ANQEPAGCKKI FGSIAFLPESPDGPASNTAPLOPQLOVFETLEITGYLISAWPDS 418
 Db 346 SNIDGFNCTKILGNLDFLITGLNVDPWHKI PALDPEKLVFRTVREITGYLNTQSPPH 405
 Qy 419 LPDLSVFQNLQVIRGRILHNGAYS--LTLOGLGISWGLRSLRELGSGLAIHHNTHLCFV 477
 Db 406 MHNFSVFNLTITIGRSLYNRGFSLLIMKNLNVTSLGFRSLKEISAGRVYISANQQLCYH 465
 Qy 478 HTVPWQDLFRNPHQALLHTA--NRPEDECVEGLACHOLCARGHCWGPQTQVNCVSOFLR 536
 Db 466 HSLNWTLLRGPSEERLDIKYDRPLGCELAEGKVCDCPLCSGGGCGWGPAGQCCLSCRYSR 525
 Qy 537 GQECVEECRVLQGLPREYVYARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPDPC 596
 Db 526 EGVCTVTCNLFQGEPRFVHEAQCFCHPECLPMEGTSTYNGSSDCAACAHFRDGPCH 585
 Qy 597 VARCPGVKPDLSYMPIWKFPDEEGACQPCPINCHTHSC--VDLDDKGPAPQASPLTSI 654
 Db 586 VNSCPHGLG--AKGPIYKYPDAQNECPCHENCTQCGNPELQDCLGQAEVLMSKPHLV 643
 Qy 655 VSAVVGILLVVLGVVFGILIKRQOKIR-KYTHRRLLQETELVEPLTPSGAMPNQAMR 713
 Db 644 IAVTVG--LAVILMILGGSFLYWRGRRIQNKRRMYRLERGESIEPLDPS--EKANKVLAR 700
 Qy 714 ILKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYV 773
 Db 701 IFKETELRKVKVLGSGVFGTVHKGIWIPGESIKIPVCIKVIEDKSRQSFQAVTDHMLA 760
 Qy 774 MAGVGSYVSRLLGICITSTVQLVTLQMPYGCCLLDHVRENRGLSGDOLLNWCMIQIAKGM 833

Db 761 VGSLDHAIHVLLGLCPGSSQLQVTOYLPGLSLDHHVQHRETLPGLLLNWGVQIAKGM 820
Qy 834 SYLEDVRLVHRLDAARNVLPSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALES 893
Db 821 YYLESHSVHRLDALRNVMKSPSOVQVADFGVADLLPPDDKQLHSEAKTPIKMALES 880
Qy 894 ILRRFRTHQSDVWYGVTVWELMTFGAKPDGIGIPAREIPDLLEKGERLPQPPICITIDVYM 953
Db 881 IHFGKYTHQSDVWYGVTVWELMTFGAEPYAGRLAETPDLLEKGERLAQPIQICITIDVYM 940
Qy 954 IMVKCMWIDSCRRFRFELVSEFMRMDPQFVVIQWEDLGPASPLDSTFYRSLLEDDDD 1013
Db 941 VMVKCMWIDENIRTFELANEFTRMADPPRYLVIKRAS-CPGTP--PAAPSVLTKE 997
Qy 1014 MGDVDAEYLVPOQGFPCDPAPGAGGMVHRRSSSTRSGGDLTLGLPSEB----- 1068
Db 998 L-----QAELEPEL-----DLDLLEABEGLATS 1023
Qy 1069 -----EAPRSLPASEG-----AGSDVPDGLGMGAAGKGLQSLPTH 1105
Db 1024 LGSALSPTGTLTRPGSQSLSPSSGYMPNMNOSLGEACLDASVLGGREGQFSRPSILH- 1082
Qy 1106 PSPQRYSEDPTVLPGETDGVY---APL-----TC-----SPOPE-----VYNOPDV 1145
Db 1083 PIPGR-----PASESSEGHVTGSEAELOEKVSVCRSRSRSPRGRDSATHSQHS 1135
Qy 1146 RPQPPSPREGP-----LPAARPAGATLERAKTLSP-GKNGVV-----KDVFAF 1187
Db 1136 LLTPVTLSPGLEEDGNGVMPDTHLGRASSREGTLSSVGLSSVLGTDEED----- 1191
Qy 1188 GGAIVENPEYLTPOGGAAPQHPHP 1210
Db 1192 ----EYEVNMRKRGSP-PRPP 1209

RESULT 11
TVFVLV

N;Contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase
C;Species: avian leukosis virus, ALV
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 11-Jun-1999
C;Accession: B00643; A00643
R;Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1985
A;Title: C-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A;Reference number: A00643; MUID:85228222; PMID:2988784
A;Accession: B00643
A;Molecule type: mRNA
A;Residues: 1-698 <N1>
A;Cross-references: GB:M10066; GB:M13881; NID:G211749; PIDN:AAA48763.1; PID:G211750
A;Note: in Genbank entry CHKERBBF, release 109.0, the source is designated as Gallus gal
C;Comment: This protein is synthesized as a gag-env-erbB protein.
C;Genetics:
A;Gene: gag-env-erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific p
F;1-6/Product: gag protein (fragment) #status predicted <GAG>
F;7-59/Product: env protein (fragment) #status predicted <ENV>
F;60-698/Product: protein-tyrosine kinase erbB #status predicted <ERB>
F;194-459/Domain: protein kinase homology <KIN>
F;202-210/Region: protein kinase ATP-binding motif
F;229/Active site: Lys #status predicted

Query Match 25.9%; Score 1766.5; DB 1; Length 698;
Best Local Similarity 52.2%; Pred. No. 1.5e-65;
Matches 374; Conservative 80; Mismatches 137; Indels 125; Gaps 18;

Qy 578 GPEADQCVACAHYKDPFCFVARCPGKVPDLSYMPIWKFPDEEGACQPCPINTCHSCVDL 637
Db 60 GP--DHCMKCAHFIDGPHCVKACPAVLGENDTL-VMKYADANAVCOLCHPNCTRGCKGP 116
Qy 638 DDGKCPAQBRASPLTSTVSNAV-VGILLVVLGVVFGVILIKRQOKIRKYVMRRLLOTEL 696

Db 117 GLEGCP---NGSKTPSTAAGVWGGLCLLVVVGIGLYLRRR-HIVRKRTLRLRLQREL 172
Qy 697 VEPLTPSGAMPNOAMRILKETELRKVKVGLSGAGFVYKGIWIPDGENVKIPVAIKVLR 756
Db 173 VEPLTPSGEAPNQAHRLILKETEFKVKVGLSGAGFVYKGLWIPGEKVKIPVAIKELR 232
Qy 757 ENTSPKANKILDEAYVMAGVSPYSRLIGICITLSTVQLVQLTMPYGCCLLDHVRNRR 816
Db 233 EATSPKANKILDEAYVMASVDNPHVCRLLGICITLSTVQLITQLTMPYGCCLLDYIREHKDN 292
Qy 817 LGSQDLNWCQIAGKMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETE 876
Db 293 IGQYLLNWCQIAGKNYLEERRLVHRDLAARNVLKTPQHVKITDFGLAKLLGADEKE 352
Qy 877 YHADGGKVPKIMMALESILRRRFTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLE 936
Db 353 YHAEGGKVPKIMMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIIPASISSVLE 412
Qy 937 KGBRLPOPPICTIDVYIMVKCMWIDSECRPRFELVSEFMRMDPQFQFVIO-NEDLG 995
Db 413 KGBRLPOPPICTIDVYIMVKCMWIDADSRPKRELTAEFKWARDPPRYLVIOQDERMH 472
Qy 996 PASPLDSTFYRSLLEDDMGDLVDAEYLVPOQGFPCDPAPGAGGMVHRRSSSTRSG 1055
Db 473 LPSPDTSKFRTLMEEEDMEDIVDADYLVPHQGF-----NSPST----- 513
Qy 1056 GGDLTGLLEPSEBAPRSP-----APSEGAGSDVFDGLGMGAAGKGLQSLPTHDPSPLO 1110
Db 514 -----SRTPLLSLSATSNNNSATNCID-----RNGQGHVREDSEFVQ 550
Qy 1111 RYSEDPTVPLPSET--DGYVAPLTCSPQPEYVNPQDVRPQPPSPREGPLPAARAGATLE 1168
Db 551 RYSDPTGNFLESIDDGFL-----PAPEYVQ--LMPKKPS----- 585
Qy 1169 RAKTLPKNGVVKDVF-----AFGAVENPEYLTPOGGAAPQHPHPAP 1213
Db 586 ----TAMVQNGIYNNISLTAISKLPMSRYQNSHSTAVDNPEYL-----NTNQSLA 633
Qy 1214 SPAPDNLXYWDQ-----DPPE-----RCAPPSTFGTPTAENPEYVLGLDVP 1254
Db 634 KTVFESSPYWQSGNHQINLDNPDYQODFLPNETKPNGLLKVPAEENPEYLRVNAAP 689

RESULT 12
TVYUHV

protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain H)
C;Species: avian erythroblastosis virus
C;Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 11-Jun-1999
C;Accession: A00644; A38022
R;Yamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.
Cell 35, 71-78, 1983
A;Title: The erbB gene of avian erythroblastosis virus is a member of the src gene family
A;Reference number: A00644; MUID:84026539; PMID:6313229
A;Accession: A00644
A;Molecule type: DNA
A;Residues: 1-604 <YAM>
A;Cross-references: GB:K01216; NID:G209676; PIDN:AAA42400.1; PID:G209678
R;Debuire, B.; Henry, C.; Benaisa, M.; Biserte, G.; Claverie, J.M.; Saule, S.; Martin, J.
Science 224, 1456-1459, 1984
A;Title: Sequencing the erba gene of avian erythroblastosis virus reveals a new type of
A;Reference number: A38022; MUID:84223957; PMID:6328658
A;Accession: A38022
A;Molecule type: DNA
A;Residues: 1-28,'W',30-139,'F',141-145,'V',147-152 <DEB>
A;Cross-references: GB:K02006
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific p
F;130-395/Domain: protein kinase homology <KIN>
F;138-146/Region: protein kinase ATP-binding motif
F;165/Active site: Lys #status predicted

Query Match 25.0%; Score 1703; DB 1; Length 604;
Best Local Similarity 52.2%; Pred. No. 5e-63;
Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 16;
QY 587 CAHYKDPFPCVPCSPGVKPLSYMPKIPDEEGACQPCPINCTHSCVDLDDKGCAPAEQ 646
DB 3 CAHFIDGPHCVKACPAVGLGENDTL-VKRYADANAVCQLCHPNCTRGCKGPLEGCP-- 58
QY 647 RASPLTSIVSAV-GILLVVVLGVVGLIKRROOKIRKYMRLLOLQETELVEPLTSGA 705
DB 59 NGSXTPSIAAGVGLCLLVVGLIGLYLRR-HIVKRTLRLLQERELVEPLTSGE 117
QY 706 MPNOAQRILKTELKVKVGLSGAGFYVYKGIWIPDGENVKI PVAIKVLRENTSPRANK 765
DB 118 APNQHURILKETEFKKVKVGLSGAGFYVYKGIWIPDGENVKI PVAIKELREATSPRANK 177
QY 766 EILDEAYVMAGVPPYSRLGICLTSTVQLVQTMPIYGLLDHVRNRRGLSGODLLNW 825
DB 178 EILDEAYVMASVDNPHVCRLLGICLTSTVQLITQIMPYGLLDVIREHKDNIGSOYLLNW 237
QY 826 CMQIAKMSYLEDVRLVHRDLAARNLVKSNHVKITDFGLARLLDDETEYHADGGKVP 885
DB 238 CVQIAKGMNLYEERLVHRDLAARNLVKTPQHVKITDFGLAKLLGADKEHYHAEGGKVP 297
QY 886 IKWMALESILRRRFTHOSDVMSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPP 945
DB 298 IKWMALESILHRIYTHQSDVMSYGVYVWELMTFGSKPYDGIIPASEISSVLEKGERLPQPP 357
QY 946 ICTIDVIMYKMWIDSECRPRELVSEFSRMDARQRFVVIQ-NEDLGPASPLDSTF 1004
DB 358 ICTIDVIMYKMWIDADSRPKFRELIARFASKMARPPRYLVIOGDERMELPSPTDSKF 417
QY 1005 YRSILLEDDGDLVDAEYLPQOGFFCPDPAPAGAGVHRRHSSSTRSGGDLTLGLE 1064
DB 418 YRTLMEEDMEDIVDAEYLPVPHOGFF-----NSPST----- 449
QY 1065 PSEEARPSPL-----APSEGAGSDVDGDLGMAAGLQSLPHDPSPLORYSEDPTVP 1119
DB 450 -----SRTPLLSLSATSNNSATNCID-----RNGQGHVREDSFVORYSDPTGN 495
QY 1120 LPSET--DGYVAPLTCSPQBYVNPQDVRPQPPSPREGPLPAARPAGATLERAKTLPFGK 1177
DB 496 FLEESIDGFL-----PAPEYVQ--LMPKKPSTAM----- 524
QY 1178 NGVVKDVPF-----GAVENPEYLTQGGAAPOHPHPPAFSPAFD 1218
DB 525 --VQNIYNFISLTAISKLPMSRYQNSHSTAVDNPEYL-----NTNQSPLAKTVFE 574
QY 1219 NLYWDDQPPPERGAPPSTFKGTPTAENPEY 1248
DB 575 SSPYWIQSNHQ-----INLDNPDY 594
RESULT 13
GOFFE
epidermal growth factor receptor - fruit fly (Drosophila melanogaster)
N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
C:Species: Drosophila melanogaster
C:Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 11-Jun-1999
C:Accession: A00640; A38021
R:Livneh, E.; Glazer, L.; Segal, D.; Schlessinger, J.; Shilo, B.Z.
Cell 40, 599-607, 1985
A:Title: The Drosophila EGF receptor gene homolog: conservation of both hormone binding
A:Reference number: A00640; MUID:85124611; PMID:2982499
A:Accession: A00640
A:Molecule type: DNA
A:Residues: 1-1330 <Liv>
A:Cross-references: EMBL:K03054
R:Wadsworth, S.C.; Vincent III, W.S.; Bilodeau-Wentworth, D.
Nature 314, 178-180, 1985
A:Title: A Drosophila genomic sequence with homology to human epidermal growth factor re
A:Reference number: A38021; MUID:85137938; PMID:2983232
A:Accession: A38021

A:Molecule type: DNA
A:Residues: 'A', 832-866, 'V', 868-943, 'QTPSLVK', <WAD>
A:Cross-references: EMBL:X02293; NID:97922; PID:CAA26157.1; PID:9929565
C:Comment: This sequence is tentative because the introns have not been identified.
C:Genetics:
A:Gene: FlyBase:Egfr
A:Cross-references: FlyBase:FBgn0003731
A:Map position: 2 57p
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
F:1-732/Domain: extracellular #status predicted <EXT>
F:733-764/Domain: transmembrane #status predicted <TM>
F:765-1330/Domain: intracellular #status predicted <INT>
F:808-1072/Domain: protein kinase homology <KIN>
F:816-824/Region: protein kinase ATP-binding motif
F:122,300,324,363,518,688,695,700/Binding site: carbohydrate (Asn) (covalent) #status pr
F:774/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:843/Active site: Lys #status predicted
F:1181/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status predict
Query Match 24.3%; Score 1654.5; DB 1; Length 1330;
Best Local Similarity 30.1%; Pred. No. 1.1e-60;
Matches 415; Conservative 176; Mismatches 417; Indels 371; Gaps 39;
QY 80 VQOYIKANSKFIGITELQRLRIVRGITOLF-----EDNYALAVLDNGDPLNNTTPTVTGASP 134
DB 38 ITNYIVIGLDLIPCTLSYRLQIIRGRTLFSLSVEEEKYALFV-----TY 81
QY 135 GGLRELQRLSLTILKGGVLIIQNPQLCYOQTLWKDIFHKNQLALTLDITNRSRACHP 194
DB 82 SKMYTLEIPDLRVLNGQVGFHNNYLNCHRTIQWSEIVSGNDAYVNYDTAPTARECPK 141
QY 195 CSPMCKSRGWSSSEDCQSLRTVTCAGGCA--RCKGPLTDCCHCCACAGCTGPKHSDC 252
DB 142 CHESCTHG--CWGEGPKNCQKFSKLTCSQPCAGRCYGFKPRECHLFCAGCTGTQKDC 200
QY 253 LACLHFNHSGICELHCPALVYNTDTPESMPNPEGRYTFGASCVTACPNYVLTSDVGSCT 312
DB 201 IACKNFDEAVSKEECPMEKYNPTTVLETNPEGYATGATCVKECP--GHLLRDNGACV 259
QY 313 LVCPLNHNEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRVAVTSANIQEPACKKIFG 372
DB 260 RSCPQDKMDKGGE-----CVPCNGPCPKTCPGVTVLH-----AGNIDSFRNCTVIDG 306
QY 373 SLAFLEPSFPG--DPASNTA-----PLOBEQLVFETLEITCYLYLISAWPDSLPLSV 424
DB 307 NIRILDQTFSGFDVYANYTMGPRIPLDPERREVSTVKEITGYLNIETHPTQFRNLSY 366
QY 425 FQNLQVIRGRILHNGAY-SITLQGLGISWGLSLRSLRSLGSLALIHHTHLCFVHTVPWD 483
DB 367 FRNLETHGRQLMESMPAALAIKVSLSYLEMRNLKQISSGVVVIQHNRLCYVSNIRWP 426
QY 484 QLFNPHQALLHTANRPEDEC-----
DB 427 AIQKEPQKVWVNVNENRADLCGRFTLISVQHNIMHIFAICREKNHLLGSVQGRLL 486
QY 505 -----
DB 487 GSWHGSVPYLQELQFQWHLHRLRLWLYIQVSINSQDKSNBHLTDACYSPSVPTSLTIER 546
QY 505 -----VGEGLA-- 510
DB 547 ARVAIQSAGLAMELEQITARSASNRHKTLPABGRQVPRVWFLGVCASARAGIAEPLAGR 606
QY 511 -----CHOLCARGHCWGPGPTQCVCNCSQFLRGQCEVEECVRLQGLPREYV---NARHCLP 562
DB 607 AVCRKCHPLCELCTNYGYHEQVCSKCTHYKRREOCETEC-----PADHYTDEQRECFQ 660
QY 563 CHPECQPNQSVTCFGEADQCVACAHYK-----DPPF-----CVARCPGVK-PDL 608
DB 661 RHPEC---NG---CTGPGADDCKSCRNFKLFDFANETGPNVSNMTNFCNCKPLMRHVNY 714
QY 609 SYMPIWKFDBEGACQPCPINCTHSCVDLDDKGPAPQASPLTSIVSAVVGILLVVVLG 668

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:25:54 ; Search time 22.9062 Seconds
(without alignments)
5267.077 Million cell updates/sec

Title: SEQ4-103-117-12
Perfect score: 6809
Sequence: 1 MEALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGDPVP 1255

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 73.*
1: PIR1.*
2: PIR2.*
3: PIR3.*
4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6730	98.8	1255	1 A24571	protein-tyrosine k
2	5957	87.5	1260	1 TVRTWU	protein-tyrosine k
3	5941.5	87.3	1254	2 I48161	p-185 precursor -
4	3160	46.4	1210	1 GQHUF	epidermal growth f
5	3130	46.0	1210	2 A53183	epidermal growth f
6	3113.5	45.7	1223	1 TVCHLV	epidermal growth f
7	3003.5	44.1	1308	2 A47253	epidermal growth f
8	2699	39.6	1166	1 S06142	protein-tyrosine k
9	2421.5	35.6	1342	2 A36223	kinase-related tra
10	2336.5	34.3	1339	2 JC4387	epidermal growth f
11	1766.5	25.9	698	1 TVFVLV	protein-tyrosine k
12	1703	25.0	604	1 TVFVLV	protein-tyrosine k
13	1654.5	24.3	1330	1 GQFFE	epidermal growth f
14	1647	24.2	544	2 S35745	protein-tyrosine k
15	1640	24.1	545	2 S00727	kinase-related tra
16	1623	23.8	540	2 B44776	protein-tyrosine k
17	1621	23.8	540	1 TVFVBB	protein-tyrosine k
18	1521	22.3	644	2 A36325	epidermal growth f
19	1311	19.3	1323	2 E88257	protein let-23 [im
20	1311	19.3	1374	2 S70712	protein-tyrosine k
21	1229	18.0	1369	2 S70713	protein-tyrosine k
22	1165	17.1	1717	1 A45558	epidermal growth f
23	1145	16.8	527	2 A42032	epidermal growth f
24	999.5	14.7	843	2 A27131	epidermal growth f
25	806.5	11.8	346	2 S13807	protein-tyrosine k
26	754.5	11.1	311	2 S13808	protein-tyrosine k
27	733	10.8	1363	2 T43220	insulin-like growt
28	720	10.6	1382	1 INHUR	insulin receptor p
29	713	10.5	1383	2 A36080	insulin receptor p

RESULT 1

A24571

protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human

N;Alternate names: c-erb-B-2 protein precursor; kinase-related transforming protein erb

C;Species: Homo sapiens (man)

C;Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999

C;Accession: A24571; A25491; A44188; B44188; I59509; I57622

R;Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T.;

Nature 319, 230-234, 1986

A;Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth

A;Reference number: A24571; MUID:86118663; PMID:3003577

A;Accession: A24571

A;Molecule type: mRNA

A;Residues: 1-1255 <YAM>

A;Cross-references: GB:X03363; NID:g31197; PIDN:CAA27060.1; PID:g31198

R;Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.

Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985

A;Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epider

A;Reference number: A25491; MUID:86016729; PMID:2995967

A;Accession: A25491

A;Molecule type: DNA

A;Residues: 737-1031 <SEM>

A;Cross-references: GB:M11767; NID:gi82163; PIDN:AAA35808.1; PID:g553282

R;Cousseus, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg,

Science 230, 1132-1139, 1985

A;Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chromo

A;Reference number: A44188; MUID:86070181; PMID:2999974

A;Accession: A44188

A;Molecule type: DNA

A;Residues: 740-910 <COU1>

A;Cross-references: GB:M12036; NID:gi83988; PIDN:AAA35978.1; PID:gi83989

A;Accession: B44188

A;Molecule type: mRNA

A;Residues: 1-517, 'RALL', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>

A;Cross-references: GB:M11730; NID:gi83986

R;King, C.R.; Kraus, M.H.; Aaronson, S.A.

Science 229, 974-976, 1985

A;Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.

A;Reference number: I59509; MUID:85272597; PMID:2992089

A;Accession: I59509

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 832-909 <REX>

A;Cross-references: GB:I29395; NID:g459807; PIDN:AAA35809.1; PID:g459808

R;Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.

Mol. Cell. Biol. 7, 2597-2601, 1987

A;Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptiona

A;Reference number: I57622; MUID:87286898; PMID:3039351

A;Accession: I57622

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-191 <TAL>

F:658-680/Domain: transmembrane #status predicted <TM>
F:723-988/Domain: protein kinase homology <KIN>
F:731-739/Region: protein kinase ATP-binding motif
F:711,191,263,535,576,634/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:691/Binding site: phosphate (Thr) (covalent) #status predicted
F:758/Active site: Lys #status predicted
F:882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 87.5%; Score 5957; DB 1; Length 1260;
Best Local Similarity 87.4%; Pred. No. 3.1e-237;
Matches 1098; Conservative 51; Mismatches 106; Indels 2; Gaps 2;

Qy 1 MELAALCRWGLLLALLPPGAASSTOVCTGDMKRLRLPASPTHLDMLRHLVGGCQVQGNL 60
Db 4 MELAAWCRWGLLLALLPPGIAAGTQVCTGDMKRLRLPASPTHLDMLRHLVGGCQVQGNL 63
Qy 61 ELTYLPTNASLSFLQDIQEVGYVLIHAHQVROVPLQRLIRVRGTQOLFEDNYALAVLDNG 120
Db 64 ELTVVPANASLSFLQDIQEVGYVLIHAHQVROVPLQRLIRVRGTQOLFEDNYALAVLDNR 123
Qy 121 DPLANNQYKANSKF-IGITELQLRSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNQL 179
Db 124 DPQNVAASTPGRTPEGLRELQRLSLTEILKGGVLIQRNPOLCYQDMVLMKDVFRKNNQL 183
Qy 180 ALTLIDNRSRACHPCSPMKGSRGWSESSDCOSLTRTVCAGGCARCKGPLPTDCCHEQ 239
Db 184 APVDIDNRSRACHPCAPACKDNHCWGESPDCCILTGITGSCARCKGRPLPTDCCHEQ 243
Qy 240 CAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRTYFGASCVTAC 299
Db 244 CAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRTYFGASCVTTC 303
Qy 300 PYNLSTDVGSCTLVCPILHNOEVTAEADGTQRCERCKSPCARVCYGLGMEHLRVRVAVTSA 359
Db 304 PYNLSTEVGSCTLVCPENNOEVTAEADGTQRCERCKSPCARVCYGLGMEHLRVRGARTSD 363
Qy 360 NIQFAGCKKIFGLAFIPESFDGDPASNTAPLOEQLOVETLEETGYLIYISAWPDSL 419
Db 364 NVQFDDGCKKIFGLAFIPESFDGDPSSGIAPLRPEQLQVETLEETGYLIYISAWPDSL 423
Qy 420 PDLVSFQNLQVIRGLIHNGAYSITLQGLISWGLRSLRELGSGLALIHNTLHCFVHT 479
Db 424 RDLVSFQNLRIIRGLIHNGAYSITLQGLIHSGLRSLRELGSGLALIHNTLHCFVHT 483
Qy 480 VPWDLFRNPHQALLHTANREDE-CVGEGLACHQLCARGHCWGPPTQCVCNCSFLRQG 538
Db 484 VPWDLFRNPHQALLHSGNREDELCVSSGLVCNLSLCAHGCWGPPTQCVCNCSFLRQG 543
Qy 539 ECVEECRVQLGPREYNARHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFPCVA 598
Db 544 ECVEECRVWKLPREYVSDKRLPCHPECPQNSSETCFGEADOCVACAHYKDSSCVA 603
Qy 599 RCPSGVPLDSYMPITWKPFPDEGACPCPINCTHSCVDLDKGCPCAPORASPLTSIVSAV 658
Db 604 RCPSGVPLDSYMPITWKPFPDEGICQPCPINCTHSCVDLDKGCPCAPORASPVTFIATV 663
Qy 659 VGILLVVLGVVFGILIKRRQKIRKYTMRLLOETELVEPLTPSGAMPNQAOMRLKET 718
Db 664 EGVLFLILVVLGVVFGILIKRRQKIRKYTMRLLOETELVEPLTPSGAMPNQAOMRLKET 723
Qy 719 ELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGV 778
Db 724 ELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGV 783
Qy 779 SPYVSRLLIGLCTSTVQLVTOLMPYGCILLDHVNRNRELGSQDILLNWCMTAKGMSYLED 838
Db 784 SPYVSRLLIGLCTSTVQLVTOLMPYGCILLDHVNRNRELGSQDILLNWCMTAKGMSYLED 843
Qy 839 VRLVHRDLAARNVLVKSPPNHVKITDFGLARLLDIDETEHADGKVPKIMWALESIILRRR 898
Db 844 VRLVHRDLAARNVLVKSPPNHVKITDFGLARLLDIDETEHADGKVPKIMWALESIILRRR 903
Qy 899 FTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVTYIMVVC 958

Db 904 FTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVTYIMVVC 963
Qy 959 WMTDSECRPFRELVSFSSWMDPQRFVVIQNEIDLGASPDLSTFYRSILLEDDMGDLV 1018
Db 964 WMTDSECRPFRELVSFSSWMDPQRFVVIQNEIDLGASPDLSTFYRSILLEDDMGDLV 1023
Qy 1019 DAEYLVPOQGFCCPDAPGAGGMVHRHRSSTSRGGDLTLGLSEPEEAPRSLAPS 1078
Db 1024 DAEYLVPOQGFCCPDPTGTAHRHRSSTSRGGDLTLGLSEPEEAPRSLAPS 1083
Qy 1079 EGAGSDVFTCDLGMGAAGKLOSLPTDPSPLORYSEDPVPLPSETGYVAPLTCSPQPE 1138
Db 1084 EGAGSDVFTCDLGMVTKGLOSLPHDLSPLORYSEDPVPLPSETGYVAPLTCSPQPE 1143
Qy 1139 YVNPQDVRFPQPPREGPLPAAPAGATLERAKTSLPGKGVKDVFAFGAVENPEVLT 1198
Db 1144 YVNPQDVRFPQPPREGPLPAAPAGATLERAKTSLPGKGVKDVFAFGAVENPEVLT 1203
Qy 1199 PQGGAAPQHPHPPAFSPAFNLVYWDQDPPERGAAPPSTFKGTPTAENPEYLGLDVVP 1255
Db 1204 PREGTASPPHPPAFSPAFNLVYWDQDPPERGAAPPSTFKGTPTAENPEYLGLDVVP 1260

RESULT 3
I48161
p-185 precursor - golden hamster
C:Species: Mesocricetus auratus (golden hamster)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999
C:Accession: I48161
R:Nakamura, T.; Ushijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishikawa
Gene 140, 251-255, 1994
A:Title: Cloning and activation of the Syrian hamster neu proto-oncogene.
A:Reference number: I48161; MUID:94193007; PMID:7908275
A:Accession: I48161
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1254 <RES>
A:Cross-references: GB:D16295; NID:G493236; PIDN:BAA03801.1; PID:G747595
C:Genetics:
A:Gene: neu
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP
F:718-983/Domain: protein kinase homology <KIN>
F:726-734/Region: protein kinase ATP-binding motif

Query Match 87.3%; Score 5941.5; DB 2; Length 1254;
Best Local Similarity 87.1%; Pred. No. 1.3e-236;
Matches 1093; Conservative 58; Mismatches 103; Indels 1; Gaps 1;

Qy 1 MELAALCRWGLLLALLPPGAASSTOVCTGDMKRLRLPASPTHLDMLRHLVGGCQVQGNL 60
Db 1 MELAAWCRWGLLLALLSPGASGTQVCTGDMKRLRLPASPTHLDIVRHLVGGCQVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIQEVGYVLIHAHQVROVPLQRLIRVRGTQOLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIQEVGYVLIHAHQVROVPLQRLIRVRGTQOLFEDNYALAVLDNR 120
Qy 121 DPLANNQYKANSKFIGITELQLRSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNQL 180
Db 121 DPLANNQYKANSKFIGITELQLRSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNQL 180
Qy 181 LTLDITNRSRACHPCSPMKGSRGWSESSDCOSLTRTVCAGGCARCKGPLPTDCCHEQ 240
Db 181 PVDIDNRSRACHPCAPACKDNHCWGESPDCCILTGITAPRAVPAARLPTDCCHEQ 240
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRTYFGASCVTAC 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRTYFGASCVTTC 300
Qy 301 YNVLSTDVGSCTLVCPILHNOEVTAEADGTQRCERCKSPCARVCYGLGMEHLRVRVAVTSA 360
Db 301 YNVLSTEVGSCTLVCPILHNOEVTAEADGTQRCERCKSPCARVCYGLGMEHLRVRVAVTSA 360

QY 361 IQEAGCKKIRGSLAFIPESFDGPPASNTAPLOEQLOVFTLEETITGYLVIASWAPSLP 420
Db 361 IQEAGCKKIRGSLAFIPESFDGPPASNTAPLOEQLOVFTLEETITGYLVIASWAPSLH 420
QY 421 DLSVFQNLQVIRGRILHNGAYSLSLTQGLGISWGLGRLSRLGSLALHFNHNLHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSLSLTQGLGIRWGLGRLSRLGSLGLVLIHRNTHLCFVHTV 480
QY 481 PWDQLFRPHQALLHTANRDEBCVGEGLACHOLCARGHCGPPTQCVNCSQFLRGQEC 540
Db 481 PWDQLFRPHQALLHSGNPEEECGLKDFACYPICAHGHCWGPPTQCVNCSHFLRGQEC 540
QY 541 VECEKVLGGLPREYVNRHCLPCHPECOPOQNGSVTCFGEADQCVACAHYKDPFPCVACR 600
Db 541 VKRECVWGLPREYVNGKHCLPCHPECOPOQNSTCTCTGSEADQCTACHYKDSFPCVACR 600
QY 601 PSGVKPDLSPYPIWKFPDEEGACQPCPINCTHSCVDLDKCPAEQASPLTISVAVVG 660
Db 601 PSGVKPDLSPYPIWKYPDEEGMCQPCPINCTHSCVDLDKCPAEQASPLTISVAVVG 660
QY 661 ILLVVVLGVGGLIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAOMRILKETEL 720
Db 661 ILLFLVIGVVVGLIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAOMRILKETEL 720
QY 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVGP 780
Db 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVGP 780
QY 781 YVSRLLGICLTSTVOLVTQMLPYGCLLDHVRNRRGLSGQDLNMCQIAKMSYLENDR 840
Db 781 YVSRLLGICLTSTVOLVTQMLPYGCLLDHVRHRRGLSGQDLNMCQIAKMSYLENDR 840
QY 841 LVHRDLAARNVLKSPNHWKLTDFGLARLLDIDETEHADGCKVPIKMALESILRRFT 900
Db 841 LVHRDLAARNVLKSPNHWKLTDFGLARLLDIDETEHADGCKVPIKMALESILRRFT 900
QY 901 HQSDVWSGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPDPPICITDVTYMTMVKCWM 960
Db 901 HQSDVWSGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPDPPICITDVTYMTMVKCWM 960
QY 961 IDSECRPRFRELVSFSEFARMADPQRFVVIQNEGLGASPLDSTFYRSLLEDMDGDLVDA 1020
Db 961 IDSECRPRFRELVSFSEFARMADPQRFVVIQNEGLGASPLDSTFYRSLLEDMDGDLVDA 1020
QY 1021 EYLVPOQGFPPDPAPGAGGVHRRSSSTRSGGDLTLGLPSEEEARPLAPSEG 1080
Db 1021 EYLVPOQGFPPDPAPGAGSTAHRHRRSSSTRSGGDLTLGLPSEEEARPLAPSEG 1080
QY 1081 AGSDVFDGDLGMAAGKQSLPTHDPSPLQRYSEDTVPPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFEGLNGATGKQSQISPRDLSPLQRYSEDTPLPTEYDGYVAPLACSPQPEYV 1140
QY 1141 NOPDVRRPQPPREGPLPAARPAGATLERAKTLSPGKNGVKDVFAPFGAVENPEYLPQP 1200
Db 1141 NOPEVRPQPLTPEGLPPVRPAGATLERPKTLSPGKNGVKDVFTEGAVENPEYLVPR 1200
QY 1201 GGAAPQHPHPPAFSAFNLYWDDPPERGAPSTFKGTATNPEYLGIDVVP 1255
Db 1201 GGSASOPH-PPALCAFDNLYWDDQDPSESGSPNPTFETTAENPEYLGIDVVP 1254

RESULT 4

GQHE

epidermal growth factor receptor precursor - human

N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB

C:Species: Homo sapiens (man)

C:Date: 15-Nov-1984 #sequence revision 27-Nov-1985 #text change 11-Jun-1999

C:Accession: A00641; A25772; S30024; A38672; A00642; A43615; A23062; A05281; A60143; A33

R:Ullrich, A.; Coussens, L.; Hayflick, J.S.; Dull, T.J.; Gray, A.; Tam, A.W.; Lee, J.; Y

rg, P.H.

Nature 309, 418-425, 1984

A:Title: Human epidermal growth factor receptor cDNA sequence and aberrant expression of

A:Reference number: A00641; MUID:84219729; PMID:6328312
A:Accession: A00641
A:Molecule type: mRNA
A:Residues: 1-1210 <LL>
A:Cross-references: EMBL:X00588; NID:g31113; PIDN:CAA25240.1; PID:g757924
A:Note: The authors translated the codon AAG for residue 540 as Asn
R:Ishii, S.; Xu, Y.; Stratton, R.H.; Roe, B.A.; Merlino, G.T.; Pastan, I.
Proc. Natl. Acad. Sci. U.S.A. 82, 4920-4924, 1985
A:Title: Characterization and sequence of the promoter region of the human epidermal gr
A:Reference number: A25772; MUID:85270438; PMID:2991899
A:Accession: A25772
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-29 <ISH>
A:Cross-references: GB:M11234; NID:g181981; PIDN:AAA53270.1; PID:g553272
R:Haley, J.; Whittle, N.; Bennett, P.; Kinchington, D.; Ullrich, A.; Waterfield, M.
Oncogene Res. 1, 375-396, 1987
A:Title: The human EGF receptor gene: structure of the 110 kb locus and identification
A:Reference number: S30024; MUID:88217333; PMID:3329716
A:Accession: S30024
A:Molecule type: DNA
A:Residues: 1-29 <HA2>
A:Cross-references: EMBL:X06370; NID:g31118; PIDN:CAA29668.1; PID:g31119
R:Haley, J.D.; Waterfield, M.D.
J. Biol. Chem. 266, 1746-1753, 1991
A:Title: Contributory effects of de Novo transcription and premature transcript termina
A:Reference number: A38672; MUID:91107677; PMID:1988448
A:Accession: A38672
A:Molecule type: DNA
A:Residues: 1-29 <HAL>
A:Cross-references: GB:M38425; NID:g181977; PIDN:AAA63171.1; PID:g553271
A:Experimental source: carcinoma cell line A431-7
R:Xu, Y.; Ishii, S.; Clark, A.J.L.; Sullivan, M.; Wilson, R.K.; Ma, D.P.; Roe, B.A.; Me
Nature 309, 806-810, 1984
A:Title: Human epidermal growth factor receptor cDNA is homologous to a variety of RNAs
A:Reference number: A00642; MUID:84245835; PMID:6330563
A:Accession: A00642
A:Molecule type: mRNA
A:Residues: 150-187 'KSVIOAV', 195, 'M', 197, 'A', 199-222, 'S', 224-304, 'RA', 307-322
'798-799, 'TD', 802-811, 'R', 813-942 <XY>
A:Experimental source: A431 human carcinoma cells, which have large numbers of EGF rece
R:Lin, C.R.; Chen, W.S.; Krueger, W.; Stolarsky, L.S.; Weber, W.; Evans, R.M.; Verma, I.
Science 224, 843-848, 1984
A:Title: Expression cloning of human EGF receptor complementary DNA: gene amplification
A:Reference number: A43615; MUID:84196372; PMID:6326261
A:Accession: A43615
A:Molecule type: mRNA
A:Residues: 713-964 <LIN>
A:Experimental source: epidermoid carcinoma cell line A431
R:Simmen, F.A.; Gope, M.L.; Schulz, T.Z.; Wright, D.A.; Carpenter, G.; O'Malley, B.W.
Biochem. Biophys. Res. Commun. 124, 125-132, 1984
A:Reference number: A23062; MUID:85046483; PMID:6093780
A:Accession: A23062
A:Molecule type: mRNA
A:Residues: 1028-1210 <SIM>
R:Weber, W.; Gill, G.N.; Speiss, J.
Science 224, 294-297, 1984
A:Reference number: A05281; MUID:84172183; PMID:6324343
A:Accession: A05281
A:Molecule type: protein
A:Residues: 25-30, 'S', 32-51, 454-467 <WEB>
R:Russo, M.W.; Lukas, T.J.; Cohen, S.; Staros, J.V.
J. Biol. Chem. 260, 5205-5208, 1985
A:Title: Identification of residues in the nucleotide binding site of the epidermal gr
A:Reference number: A60143; MUID:85182650; PMID:2985580
A:Accession: A60143
A:Molecule type: protein
A:Residues: 740-744, 'X', 746-747 <RUS>
R:Motczkowski, B.; Mosig, G.; Cohen, S.
Nature 309, 270-273, 1984
A:Title: ATP-stimulated interaction between epidermal growth factor receptor and superc
A:Reference number: A38023; MUID:84191554; PMID:6325948
A:Contents: annotation; receptor activity

A;Molecule type: mRNA
A;Residues: 969-971,'K',973-1115,'D' <EIS>
A;Cross-references: EMBL:Z12608
R;Heiseremann, G.J.; Gill, G.N.
J.; Biol. Chem. 263, 13152-13158, 1988
A;Title: Epidermal growth factor receptor threonine and serine residues phosphorylated i
A;Reference number: A28941; MUID:8830814; PMID:3138233
A;Accession: A28941
A;Molecule type: protein
A;Residues: 689-694,'K',696-704,'L',706-707;989-992,'XX',995-996,'X',998-1000;1002-1009,
R;Hibbs, M.L.; Dunn, A.R.; Alexander, W.S.
submitted to the EMBL Data Library, April 1994
A;Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor
A;Reference number: S45325
A;Accession: S45325
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-971,'K',973-1210 <VER>
A;Cross-references: EMBL:X78987; NID:G488830; PIDN:CAA55587.1; PID:G488831
R;Paria, B.C.; Das, S.K.; Andrews, G.K.; Dev, S.K.
Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993
A;Title: Expression of the epidermal growth factor receptor gene is regulated in mouse b
A;Reference number: I49643; MUID:93126380; PMID:7678348
A;Accession: I49643
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 12-20;22-132 <RES>
A;Cross-references: GB:L06864; NID:G193001; PIDN:AAA53029.1; PID:G567201
C;Genetics:
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphopro
F;1-24/Domain: signal sequence #status predicted <SIG>
F;648-670/Domain: transmembrane #status predicted <TM>
F;712-977/Domain: protein kinase homology <KIN>
F;720-728/Region: protein kinase ATP-binding motif
F;680,695/Binding site: phosphate (Thr) (covalent) #status experimental
F;697,1070,1071/Binding site: phosphate (Ser) (covalent) #status experimental
F;993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental
F;1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental
F;1197/Binding site: phosphate (Tyr) (covalent) #status experimental

Query Match 46.0%; Score 3130; DB 2; Length 1210;
Best Local Similarity 49.5%; Pred. No. 2.5e-121;
Matches 630; Conservative 171; Mismatches 361; Indels 110; Gaps 23;

Qy 11 LLLALLPPGAA--STQVCTGDMKRLRSPASETHLDMRLHLYQCGVQVQGNLEITYLPTN 68
Db 14 LLTALCAAGALEEKVKVCGQTSNRLTQLGTPEDFHLSLQRMYNVNCVVLGNLEITYVQRN 73

Qy 69 ASLSPLQDIOEVQGVLTAAHQVROVPLORLRIVRGTOLFEDNYALAVLDNGDPLNNQYI 128
Db 74 YDLSFKLTQIEVAGVYLTALNTVERIPLNQLIKNALYENTYALILSN----- 124

Qy 129 KANSKFIQITELQLASLTLEILKGGVLIQRNPOLCYQDTILWKDI----FHKNNLALPLI 184
Db 125 -YGTNRRTGLRELPMENLQELLIGAVRFSNNPILCNMDTIQWRDIVQNVFMSMDL--- 180

Qy 195 DTNRSRACHPCSPMKCGSRGWESSEDCQSLTRTVACGGCA-RCKGPIPTCCCHQCAAG 243
Db 181 -QSHFSSCPKCDPSCPNGSCWGGGEGNCKLTKIAQOCCSHRCGRSPSCCHNOCAAG 239

Qy 244 CTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFEGASCVTACPNY 303
Db 240 CTGPRESCLVQKQFQDQATCKDTPPLMLNPTTYQMDVNPGEKISGATCVKCKPRNY 299

Qy 304 LSTDVGSCTVCLPLHNQVTAEDGTQCEKSKPCARVCYGLGMEHLREVRVAITSANTQE 363
Db 300 VVTDHGSVCVRAGPDYIEV-EEDGIRKCKKDGCPKRCVNGIGIGEPKDKTILSINATNIK 358

Qy 364 FAGCKKIFGSLAFIPESFDGDPASNTAPLOEQVFTLEITTYLYVISAWPDSLPDLIS 423
Db 359 -FKYCTAISGDLHILPVAFGKGSFTRTPELDPRELEILKTVKEITGFLLIQAWPDNWTDLH 418

RESULT 6

TVCHLV

epidermal growth factor receptor precursor - chicken

N;Contains: protein-tyrosine kinase (SC 2.7.1.112) erbB

C;Species: Gallus gallus (chicken)

C;Date: 28-Feb-1986 #sequence_revision 05-May-1995 #text_change 04-Feb-2000

C;Accession: A27720; A00643

R;Lax, I.; Johnson, A.; Howk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Vennet

Mol. Cell. Biol. 8, 1970-1978, 1988

A;Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mo

A;Reference number: A27720; MUID:88261272; PMID:3260329

Qy 424 VFONLQVIRGRIILHNGAYSLTLOGLGISWLGSLRLSRLSGSLALIHHTHLCFVHTVPWD 483
Db 419 AFENLEIIRGTRKOHGQFSLAVVGLNITSLGRLSKLEISDGDVLIISGNRLCYANTINWK 478

Qy 484 QLFNRPHQALLHTANRPEDSCVGEGLACHOLCARGHCWGPGTQCVNCSPFLRQCEVVE 543
Db 479 KLFGTPNQKTKIMNRAEKDKAVNHVCNPLCSSEGCVGPEPRDCVSCQNVSRGECVEK 538

Qy 544 CRVLOGLPREVYNARHCLPCHPCEQOPQNGSVTCFGEADOCVACAHYKDPFPCVAPCSG 603
Db 539 CNILEGEPREFVENSECICQHPCEPLQAMNITCTGRPDNCICQAHYIDGPHCVKTCPCAG 598

Qy 604 VKPDLSPYMPKPFDEBGAQCPINCTHSCVDLDKGCAPAEORASPLTSTVSAAVVGILL 663
Db 599 IMGENNTL-VVKYADANNVCHLCHANCTYGCAGPCLQCGCEVWSPGKIPSIATGIVGGLL 657

Qy 664 VVVLGVVFGI-LIKRROOKIRKYTMRLLOETELVEPLTPSGAMPNQAQWRIKETEURLK 722
Db 658 FIVV-VALGIGLPMRRRHIVRKRTLRLQLQERELVEPLTPSGEAPNQAHLRLKETETPKK 716

Qy 723 VKVLGSAFCGVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMAGVSPYV 782
Db 717 IKVLGSAFCGVYKGLWIPGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPHV 776

Qy 783 SRLIGICLTSTVOLVTOLMPYVGLLDHVRNRRGLGSODLLNWCMIQAKMSYLEDEVRLV 842
Db 777 CRLLIGICLTSTVOLVTOLMPYVGLLDYVREHKONIGSQYLLNWCVAQAKGNYLEDEDRVL 836

Qy 843 HRDLAARNVLKPNHVKITDFGLARLLDIDETEHADGGKVPKIKMMALESILRRRTHQ 902
Db 837 HRDLAARNVLKTPQHVKITDFGLAKLGAEEYHAEGGKVPKIKMMALESILHRIVTHQ 896

Qy 903 SDVMSYGVTVWELMTFCAKYDGIIPAREIPDLLEKGERLPQPPCTTIDVYIMVYKCMWID 962
Db 897 SDVMSYGVTVWELMTFGSKPYDGIIPASDISILEKGERLPQPPCTTIDVYIMVYKCMWID 956

Qy 963 SECRPRELVSFSPSRMARDPQFVWITQ-NEDLGPASPLDSTFYRSLLEDDDDMDGLVDAE 1021
Db 957 ADSRPFRELLIFESQWARPQRYLVLOGDERMHLPSPTDSNFYALMDEDEMDVVDAD 1016

Qy 1022 EYLVPQGGFFCPDPPAGAGGMVHRRSSSTRSGGDLTLGLPSEBEAPRSLAPSEGA 1081
Db 1017 EYLVPQGGFF-----NSPST-----SRTPLLSLSA 1042

Qy 1082 GSDVFDGLGGAAGKGLQSLPTHDPSPLOYSSEDPVPLPSET--DCYVAPLTCSPQPEY 1139
Db 1043 TSN-----NSTVACINRNGSCRVKEDAFQRYSSDPTGAVTEDNIDDAFL-----PVPEY 1092

Qy 1140 VNQPDVVRQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVYKDVFAFGGAVENPEYL-T 1198
Db 1093 VNQ-SVPRKPAQSVQNVYHNQPLHP-----AFGRDLHYQN--PHSNVAGNPEYLNT 1141

Qy 1199 PQGGAAPQPHPPAPFAPDNLYYWDQ-----DP-----PERGAPPSTFKGTPT 1242
Db 1142 AQ-----PTCLSSGFNSPALWIKQSHQMSLDNPDYQODFFPKETKPNGIFKG-PT 1191

Qy 1243 AENPEYLGLDVP 1254
Db 1192 AENAEYLRVAPP 1203

A:Accession: A27720
A:Molecule type: mRNA
A:Residues: 1-1223 <LAX>
A:Cross-references: GB:M20386
R:Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1985
A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A:Reference number: A00643; MUID:85228222; PMID:2988784
A:Accession: A00643
A:Molecule type: mRNA
A:Residues: 585-1223 <NIL>
A:Cross-references: GB:M10066
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
C:Specific protein kinase
F:1-30/Domain: signal sequence #status predicted <SIG>
F:31-1223/Product: epidermal growth factor receptor #status predicted <MAT>
F:31-654/Domain: extracellular #status predicted <EXT>
F:81-307/Domain: EGF receptor extracellular domain repeat <EE1>
F:397-610/Domain: EGF receptor extracellular domain repeat <EE2>
F:655-677/Domain: transmembrane #status predicted <TM>
F:678-1223/Domain: intracellular #status predicted <INT>
F:719-984/Domain: protein kinase homology <KIN>
F:727-735/Region: protein kinase ATP-binding motif
F:136,202,361,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent) #s
F:192,650/Binding site: carbohydrate (Ser) (covalent) #status predicted
F:687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:754/Active site: Lys #status predicted
F:1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat

Query Match 45.7%; Score 3113.5; DB 1; Length 1223;
Best Local Similarity 48.5%; Pred. No. 1.2e-120;
Matches 629; Conservative 175; Mismatches 348; Indels 145; Gaps 24;

Qy 8 RWGLLLALLPPGAA-----STQVCTGDMKRLRLPASPETHLMLRLHYQCCQVQGNLE 61
Db 13 RGAALVLLLLGVALCSAVEKEKVCQGTNNKLTQLGHVEDFTSLQRMNNECVVLSNLE 72
Qy 62 LTYLPTNASLSFLDIOEVOQYVLIANNQVQVPLRLRIVRGTLQEDNEDYALAVLNGD 121
Db 73 ITVYEHNRDLTKTIQEVAGYVLIANNMVDVPLENQLIIRGNVNLVDNSPALAVLSNYH 132
Qy 122 PLNNQYIKANSKFICITELQRLSRLTEILKGGVLIRPOLCYQDTILMKDIFHKNLAL 181
Db 133 MNKTK-----GLRELPMKRLSEILNGGVKLSNNPKLNCMDVTLNNDIIDSRL-PL 182
Qy 182 TLID-THRSRACHPCSPMKSGRCWGSSEDCQSLTRITVCAGCA-RCKGLPLPTDCHEQ 239
Db 183 TVLDFASNLSSCPKCHPNCTEDHCWGAGEQNCQTLLTKVCAQCCGRCGRKVPSCCHNQ 242
Qy 240 CAACCTGPKSDCLACLHNSGICELHCPALVYNTDTPEMNPBGRYTFGASCYTAC 299
Db 243 CAACCTGPRSDCLACRKRFRDADTKCTCPPLVLYNPTYQMDVNPBGKYSFGATCVR 302
Qy 300 PYNLTSDVGSCTLVCLPHNOEVTAEQTCERCKSPCARVCYGLGWEHLREVRVTS 359
Db 303 PHNVVTDHGSCVPSNCTDTEYV-EENGVRCKKCDGLCKSKVNGIGIGELKGLSINAT 361
Qy 360 NIOBFAGCKKIFGSLAFPLPSFDGDPASNTAPLOEQVLPETLEEITGYLITSAPDLS 419
Db 362 NIDSFKNCTKINGDVSILPVAFLGDAFTKTLPLDPKXLDVPRTVKEISGFLLIQAWPDNA 421
Qy 420 PDLVSFQNLQVIRGILHNGAYSITLQGLGHSWGLSLRSLGSLALIHNTLCFVHT 479
Db 422 TDLVAFENLEIRTKHQGOYSLAVNLKIQSLGLRSLKEISDGDITAMKNKMLCYADT 481
Qy 480 VPWDLFRNPHQALLHTANRPEDECVGEGLAGLACHOLCARHGCHGPGPTQVCNCSOFLRGOE 539
Db 482 MNWRLSFLATQSOKTKIITQRNKNDCTADRHVCDPLCSDVGCWGPFPCHFCRFFSQQE 541
Qy 540 CVEECRVLOGLPREYVNVNARHCLPCHPBCQPQNG---SVTCFGPEADQCVACAHYKDPFPC 596

Db 542 CVKQCNTLQGEPRFERDSCFLPCHSECLVQNSTAYNTTCSGPDCHMKCAHFIDGPHC 601
Qy 597 VARCPSPGVKPDLSVMPPIWKPEDEEGACQCPINCTHSCVDLDDKGCPCAEQASPLTISVS 656
Db 602 VKACPAGVIGENDTL-VWKYADANAVCOLCHPNCCTCKGPGLEGCP---NGSKTPSIAA 657
Qy 657 AVV-GILLVVLGVVFGILIKRRQKIRKYTMRLLOETELVBEPLTPSGAMPNOAQMRIL 715
Db 658 GWVGLLCLVVGGLIGLYLRRR-HIVRKETLRLQLQERELVBEPLTPSGEAPNQAHRLIL 716
Qy 716 KETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDVYMA 775
Db 717 KETEFKVKVYLGSGAGFTVYKGLWIPGEKVKIPVAIKELREATSPKANKEILDVYMA 776
Qy 776 GVSPYVSRLLGLTSTVQLTQMPYGLLDHVRNRCGLSODLNNCMQIAKMSY 835
Db 777 SVDPHVCRLGLTSTVQLITQMPYGLLDYIREHKONIGSOYLLNMCVQIAKMN 836
Qy 836 LEDVRLVHRLDAAARNVVKSPNHVKITDFGLARLLDIDETEHADGKVPKMMALESIL 895
Db 837 LEERRLVHRLDAAARNVVKTPQHVKITDFGLAKLGADEKEYHAEGKVPKMMALESIL 896
Qy 896 RRRFTHOSDWSGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPICITIDVYMM 955
Db 897 HRIYTHQSDVMSYGVTVWELMTFGSKPYDGPASEISSVLEKGERLPQPPICITIDVYMM 956
Qy 956 VKCMMIDSECRPRELVSFSSRMARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDDDM 1014
Db 957 VKCMMIDADSRKPRELIAEFSEKWARDPPRYLVIGQDERMHLSPPTDSKYRITLMEEDM 1016
Qy 1015 GDVLDAEYLVPOQGFPCDPAPGAGGMVHRRSSSTRSGGDLTLGLSPSEAEAPRSP 1074
Db 1017 EDIVDADEYLVPHQGF-----NSPST-----SRTPT 1042
Qy 1075 L-----APSGAGSDVFDGDLGMCAKGLQSLTHDPSPLOQVSEDPVPLPSET--DGY 1127
Db 1043 LLSLSLATSNNSSATNCID-----RNGQGHVREDSPVQYSSDPTGNFLSEIDDDG 1094
Qy 1128 VAPITCSQPEYVNPDPVRPPSPREGPLPAARAGATLERAKTLSPGKNGVVKDFV- 1185
Db 1095 L-----PAPEYVQ--LMPKPS-----TAMVQNIYNNISLT 1125
Qy 1186 -----AFGAVENPEYLTPOGGAAPQHPPPAFSPAFNLYYWDQ----- 1225
Db 1126 AISKPLPMSRYQNSHSTAVDNPEYL-----NTQSPKATVFPSSPYIQSGNHQIN 1177
Qy 1226 -DPPE-----RGAPPSTFKGPTTAENPEYLGIDVP 1254
Db 1178 LDNPDYQQDFLPNETKPNGLLVKVPAAENPEYLRVAAP 1214

RESULT 7

A47253

epidermal growth factor receptor, HER4 - human

C:Species: Homo sapiens (man)

C>Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999

C:Accession: A47253

R:Plowman, G.D.; Culouscou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.; N

Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993

A:Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epidermal

A:Reference number: A47253; MUID:93189574; PMID:8383326

A:Accession: A47253

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: nucleic acid

A:Residues: 1-1308 cDPO.

A:Cross-references: GB:I07868; NID:G337359; PIDN:AAB59446.1; PID:G337360

A>Note: sequence extracted from NCBI backbone (NCBI:P126942)

C:Superfamily: epidermal growth factor receptor; protein kinase homology

C:Keywords: ATP; growth factor receptor

F:716-981/Domain: protein kinase homology <KIN>

F:724-732/Region: protein kinase ATP-binding motif

Db 123 -----YQKNPPSDVTVQVGLKQLQLNLNLTLLSGGVKVSHPNLLCNVETINWMDIVDK 175
Qy 176 NNQALTLIDNRRACHPCSPMCKSGRCWGESSEDCQSLTRTVCAAGC-ARCKGPLPTD 234
Db 176 TSNPTMNLIPHAERQCKDHGCVNGSCWAPGPGCHQKFTKLLCAEQCNERCGRPKPID 235
Qy 235 CCHQCAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRTYTGAS 294
Db 236 CCNEHCAGGCTGPRATDCLACRDNDDGTCKDTCPPPKIYDIVSHQVVDNPNIKYTGAA 295
Qy 295 CVTACPVNLTSDVSGCTLCVPLHNOEVTAEADGTORCEKSKPCARVCYGLGMEHLREVR 354
Db 296 CVKECPNYYVTE-GACVRSACSAGMLEVD-ENGKRSKPCDGVCPKCDGIGISLNTI 353
Qy 355 AVTSANTIOEAGCKKIFGLAFLESFQGDPASTAPLQPEQLQVFTLEETITGYLISA 414
Db 354 AVNSTNIRSPNCTKINGDIILNRNSFEGDPHYKIGTMDPEHLNLTIVKEITGYLIMW 413
Qy 415 WPSLPLDSVQNLOVIRGRILHNGAYS-LTLOGLGISWLGRLSRLGSLGLAIHINTH 473
Db 414 WPNMTSLSVFQNLLEIRGRTTFSRGFSFVVQVRHLQWLGLRLSRLKEYSAGNVILKNTLQ 473
Qy 474 LCFVHTVPWDLFRNPHQALLHTANRPEDECVGSLACHQLCARGHGWGPGTQCVNCSQ 533
Db 474 LRYANTINWRRLFRSEDSISYDART-----ENQTCNNECSBDGCM-PGPTMCVCSLH 525
Qy 534 FLRGOECVEECRLQGLPREYVNAHCLPCHPECPQNGSVTCFPGPEADQCACAHYKDP 593
Db 526 VDRGRCVASCNLLQGEPREAQVDCRCVQCHQECVCLVQDTSLTCTGPGPANCSSAHFQDG 585
Qy 594 PFCVARCPGKVPDLSYMPIWKFDEBAGCQPCPINCTHSCVDLDDKGCAPAEQASPLTS 653
Db 586 PQICPRCPHGTLGSDGDTL-INKYADKMGCCQPCQNCCTQGCGSLGCRGD-IVSHSSL 643
Qy 654 IVSAVVGILLVVLGVVFGILLKROOKIRKYTWRRLLQETELVEPLTPSGAMPNQOMR 713
Db 644 AVGLVGLLITIVALLIVLVLVRLRRRIK-RRKTRICLLQEKELVEPLTPSGQANQAPLR 702
Qy 714 ILKETELRKVKVLGSGAGFTVYKGIWIPDGNVNKIPVAIKVLRNTPSPKANKEILDYAV 773
Db 703 ILKETEFKDRVLGSGAGFTVYKGLNPDGENIRIPVAIKVLRATSPKVNQEVLDYAV 762
Qy 774 MAGVGSPPVSRLLGLCTSTVOLVTQLMPYQCLLDHVRNGLSGQDLNWCQIAKGM 833
Db 763 MASVDHPHVCRLGLCTLSAQLVTQLMPYQCLLDYVRQHERICGQWLLNWCQIAKGM 822
Qy 834 SYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHYAGGKVPDKWMALES 893
Db 823 NYLBERHLVHRDLAARNVLLKNPNHVKITDFGLSKLLTADKEYOAGDGKVPDKWMALES 882
Qy 894 ILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYM 953
Db 883 ILQWYTHQSDVMSYGVTVWELMTFGSKPYDGIIPAKEIASVLENGERLPQPICTIEVYM 942
Qy 954 IMVKCWMIDSCRPRFRELSEFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDD 1013
Db 943 IILKCWMIDPSRPRFRELVEGFQMDRPSYLVIOG--NLPSLSDRRLFSRLSSDD 999
Qy 1014 MGLDVAEEYLVPOGFCPPDPAFAGAGGVHHRSSSTRSGGDLTLGLEPSEEAAPS 1073
Db 1000 --DVVDADEYLLPYKRI-----NRQGS-----E 1020
Qy 1074 PLAPSEAGSDVDFGDLGMGAAGKGLQSLTPDHPLOQRYSDPTV-PLPSETDGYVAPLT 1132
Db 1021 PCIEPTGH-----PVRENSITLNIISDPTQNALEKDLGDH----- 1055
Qy 1133 CSPOEYVNPQDVRPQP-----PSPRE-----GPLP-AARPAGATLERAKTLPQKNG 1179
Db 1056 -----EYVNPQGETSSRLSDIYNPNVEDLTDGMPVSLSSQEAETNFSRPEYLNQNS 1110
Qy 1180 VKDVFAGGAVENPEYLTPOGGAAPQHPPPAPSPADNLYYWDQPPERGAPSTFKG 1239
Db 1111 L---PLVSSGSMDDPEY---QAG-----YQAAF-----LPQTGALTNGMF 1145

Qy 1240 TPTAENPEYLG 1250
Db 1146 LPAENLEYLG 1156

RESULT 9

A36223
K:Species: Homo sapiens (man)
C:Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000
C:Accession: A36223; I59164
R:Kraus, M.H.; Ising, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.
Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989
A:Title: Isolation and characterization of ERBB3, a third member of the ERBB/epidermal
A:Reference number: A36223; MUID:90083234; PMID:2687875
A:Accession: A36223
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1342 <KRA>
A:Cross-references: GB:M29366
R:Plowman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todaro, G.
Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990
A:Title: Molecular cloning and expression of another epidermal growth factor receptor-r
A:Reference number: I59164; MUID:90311312; PMID:2164210
A:Accession: I59164
A:Status: preliminary; translated from GB/EMBL/DBDJB
A:Molecule type: mRNA
A:Residues: 1-559, 'G', 561-957, 'F', 959-1063, 'G', 1065-1342 <RES>
A:Cross-references: GB:M34309; NID:9183990; PID:AAA35979.1; PID:g306841
C:Genetics:
A:Gene: GDB:ERBB3; HER3
A:Cross-references: GDB:119880; OMIM:190151
A:Map position: 12q13-12q13
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homol
C:Keywords: Atp; phosphotransferase
F:707-972/Domain: protein kinase homology <KIN>
F:715-723/Region: protein kinase ATP-binding motif

Query Match 35.6%; Score 2421.5; DB 2; Length 1342;
Best Local Similarity 40.4%; Pred. No. 3.1e-92;
Matches 530; Conservative 192; Mismatches 460; Indels 129; Gaps 32;
Qy 10 GLLIALLPPGAA--STQVCTGTDMLKRLPASPTHLDMLRHLRYGCGCQVQGNLELTPLT 67
Db 11 GLFLSLARGSEVGNSSQAVCPGTLNGLSVTGDENQVQTLKLYERCEVVMGNLEIVLTGH 70
Qy 68 NASLSFLQDIOEQVGVYLIHNNQVRQVPLQRLRIVRGTQQLFEDNYALAVLDNGDPLNNQY 127
Db 71 NADLSFLQWIREVTGYVLVAMNBFSTLPLNLRVVRGTQVYDGKFAIFVMLN----- 122
Qy 128 IKANSKFIGITELQLRSLTEILKGGVLIQNPOLCQVDTTLWKDIFHKNNQLALTLDTN 187
Db 123 YNTNSSH-AURQLRLTQLTEILSGGVYIEKDKLCHMDTIDWRDIVDRD---AEIVVKD 178
Qy 188 RSRACHPCSPMCKSGRCWGESSEDCQSLTRTVCAAGC-ARCKGPLPTDCHEQCAAGCTG 246
Db 179 NGRSCPPCHEVCKG-RCWGPSEDCQTLTKTIAPQCNGHCFGPNPNQCCHDECAGCGSG 237
Qy 247 PKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRTYTGASCTVACPYNLYST 306
Db 238 PQDITDCFACRHFNDGACVPCPQLPVNKLITQLEPNPHTKYQYGGCVASCAPHNEV-V 296
Qy 307 DVSGCTLVCLPHNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRATYSANIQFAG 366
Db 297 DOTSCVRACPPDKMEVD-KNGLKMCBPCGGLCPKACEGTGSG--SRFTQTVDSNIDFVN 353
Qy 367 CKKIFGSLATLPESFGDGPASNTAPLQPEQLQVFTLEETITGYLISAWPDSLPLDLSVFQ 426
Db 354 CTILGNLDFLITLNGDPPHHKIPALDPEKLVNFRITVREITGYLNTQSPHMHNFVSFS 413
Qy 427 NLQVIRGRILHNGAYS-LTLOGLGISWLGRLSRLGSLGLAIHINTHLCFVHTVPMDQL 485

Db 414 NLTTIGRSLYNRGFSLLIMKLNVTSLGFRSLKEISAGRIYISANRQLCYHHSLNWKV 473
Qy 486 FRNPHQALLHTA-NRPEDECVGEGLACHOLCARGHCWGPGTQCVCNCSQFLRGQECVEEC 544
Db 474 LRGPTEERLDIKHNRPRDCVAEGKCDPLCSSGGCGWPGGQCLSCRNRYRGVGVVTHC 533
Qy 545 RVLOQLPREYVNAHCLPCHPECPONGSVTCFPEADQCACAHYKDPFPCVACRCPGV 604
Db 534 NFLNGEPREFAHEAECFSCHEPCQMEGTATCNGSGSDTCAQCAHFRDGPCHVSCPHGV 593
Qy 605 KPDLSYMPIWKPPEBEGACQCPNCNTHSCVDLDDKGCAPAQRA-----SPLTSTIVSAVVG 660
Db 594 LG--AKGPIYKYPDVQNECRPCHENCCTQCKGPELQDCLGOTLVILGTHLTMTALVIAG 651
Qy 661 ILLVVVLGVVGGILIKRQOKIR-KYTHRRLLQETELVEPLTPSGAMPNQAMRLKETE 719
Db 652 --LVVIFMVLGGTFLYWRGRIQNKRAMRYLERGESIEPLDPS--EKANKVLARIFKETE 708
Qy 720 LRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVGS 779
Db 709 LRKLVKLGSGVGTGVHKGWVLPAGESIKIPVCIKVIEDKSGRQSFQAVTDHMLAIGSLDH 768
Qy 780 PYVGRLLGICLTSTVQLVTLQMPYGCLLDHRVNRGRGLSQDLNMCWQIAKGSYLEDV 839
Db 769 AHIVRLGLCPGSSQLVQYLPGLSLDHRVQRHGRALGPOLLNMGVQIAKGMYYLEH 828
Qy 840 RLVRDLAARNVLKSPNHKIDPGLARLDDIDETEHADGGKVPKIKWMALESLRRF 899
Db 829 GMVHRNLAAARNVLLKSPSQVADFGVADLLPPDDKQLLYSEAKTPIKMALESLHFQY 888
Qy 900 THQSDVMSYGVTVWELMTFGAKPYDGIPIARIIPDLLEKGERLPOPICTIDVYIMVKCW 959
Db 889 THQSDVMSYGVTVWELMTFGAEPYAGLRLAEVDPDLLEKGERLAQPICTIDVYIMVKCW 948
Qy 960 MIDSCRRPRRELVESEFMRDPORFVVIQNELDGPA---SPLDSTFYRSLLEDDMD 1016
Db 949 MIDENIRPTEKLANEFTRMARDPRYLVIKRES--GPGIAPGPEPHGLTNKLEBELEP 1007
Qy 1017 LVDAEYILVPOGFCPPAPCAGGVHHRSSSTRSGGDLTLGLFP--SEEAAPRPL 1075
Db 1008 ELDDLDELEBED-----NLAATTLGSALESFVGTTLNRPGRSQSL 1048
Qy 1076 APSEGAGSDVDGDLGMAAKQLSLPTH--PSPLQRYSEDPVPLP-----SETDGYV 1128
Db 1049 SPSSGY-MPMQGNLQESQESAVSGSERCPVSLH-----PMPRGLASSESGHV 1101
Qy 1129 A-----PLTCSQPE-----YVNPQDVRPQPPSPREGP-----L 1157
Db 1102 TGSEAELOEKVSMCRSRSRSPRPGDSAYHSQRHSLLTPTVTPLSPPGLEEDVNGYVM 1161
Qy 1158 PAARPAGATLBRACKTLP--GKNGVV-----KDVFAFGAVENPEYLTPOGGAAPQHP 1210
Db 1162 PDTHLKGTPTSSREGTLSSVGLSSVLGTBEED-----EVEYEMNRRRHSP-PHP 1212
Qy 1211 PAFSPAFDNLVYWD-----QPPPERGAPPSTFKGPTAENPEYL 1249
Db 1213 RPSLSLEELGYEYMDVGDLSASLSTQSCPLHPVIMPTAGTTPDEDIYEM 1263

RESULT 10
JC4387

epidermal growth factor receptor homolog precursor - rat
N:Alternate names: ErbB3 protein; HER3 protein
C:Species: Rattus norvegicus (Norway rat)
C:Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998
C:Accession: JC4387
R:Hellyer, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.
Gene 165, 279-284, 1995
A:Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protein.
A:Reference number: JC4387; MUID: 96096535; PMID: 8522190
A:Accession: JC4387
A:Molecule type: mRNA
A:Residues: 1-1339 <HEL>

A:Cross-references: GB:U29339; NID:g915389; PID:g915390
A:Experimental source: liver
A:Note: The authors translated the codon AAC for residue 369 as Thr and GTT for residue
C:Comment: This protein is a functional heregulin receptor that transduces signals to
C:Genetics:
A:Gene: ErbB3
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homol
C:Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane protein
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1339/Product: epidermal growth factor homolog #status predicted <MAT>
F:640-659/Domain: transmembrane #status predicted <TM>
F:705-970/Domain: protein kinase homology <KIN>
F:713-721/Region: protein kinase ATP-binding motif
F:939,1051,1156,1194,1196,1219,1257,1259,1273,1286,1325/Binding site: phosphate (Tyr)

Query Match 34.3%; Score 2336.5; DB 2; Length 1339;
Best Local Similarity 40.5%; Pred. No. 9.4e-89;
Matches 520; Conservative 172; Mismatches 436; Indels 155; Gaps 34;

Qy 3 LAALCRWGLLLALLPPGAA---STQVCTGTDMLRLPASPEHLDMRLHLYQGVVQGN 59
Db 7 LQVLC-----FLLSLARSEMNSQAVCPGTILNGLSVTGADNQYQTLYLYKECEVWGN 62
Qy 60 LELTYLPTNASLFLQDIQEVQGYLVIAHNOVRQVPLQRLIRIVRGTOQLFEDNYALAVLDN 119
Db 63 LEIVLTCHNADLSFLQWIREVTAIVLVAMNEFSLPLNLRVVRGTQVYDGKFAIFVMLN 122
Qy 120 GDPLNNQYIKANSKFIGITELQLRSLTEILKGGVLIOENPOLCYQDTILWKDIFHKNNQL 179
Db 123 -----YNTNSSH--ALRQKFTQLTEILSGGVYIEKNDKLCHMDTIDRDIVRVR--- 170
Qy 180 ALTLDITNRRACHPCSPMKGSRGWESSEDCSLTRTVTCAGC--ARCKGPLTDCCHE 238
Db 171 GAELVKNKANCPCPCHEVCCKG--RCWGPDPDDCOILTKTICAPQNCRCFCGPNQCCHD 229
Qy 239 QCAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTNTDTFESMPNPEGRYTFGASCVTA 298
Db 230 ECAGGCGSPQDTCDFACRRFNDSGACVPCPEPLVYNKLTFLQEPNPHTKYQYGGCVAS 289
Qy 299 CPNYLSTDVSGCTLVCPHLNQEVTAEDGTQRCCKSPCARVCYGLMEHLREVAVTS 358
Db 290 CPHNFV--VDQTFVCVRACPPDKMEVD--KHGLKMCPECGGLCPKACEGTSG--SRYTQVDS 345
Qy 359 ANIQEAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFELEETIGLYISAMPDS 418
Db 346 SNIDGFVNCYKILGNLDFLITGLNVDPHWKIPALDPEKLNVRVTRIVRITGYLNTQSPHP 405
Qy 419 LPDLSVFQNLQVIRGRILHNGAYS--LTLQGLGISWLGLRSLRELGLGLLIHHTHLICFV 477
Db 406 MHNFSVFSNLTTIGRSLYNRGFSLLIMKLNVTSLGFRSLKEISAGRIYISANRQLCYH 465
Qy 478 HTVPWDOLFNRPHQALLHTA--NRPEDECVGEGLACHOLCARGHCWGPGTQCVCNCSQFLR 536
Db 466 HSLNWTLLRGPSEERLDIKYDRPLGECLAEGKVCDFLCSSGGCGWPAQGCCLSCRNYSR 525
Qy 537 GQECVEECRVQLGUPREYVNAHCLPCHPECPONGSVTCFPEADQCACAHYKDPFPC 596
Db 526 EGVCTHCFNFGQEPREFVHEAQCFSCHEPCLPMEGISTYNGSSDACARCAHFRDGPCH 585
Qy 597 VARPSPGVKPDLSYMPIWKPPEBEGACQCPNCNTHSC--VDLDDKCPAQEASPLTISI 654
Db 586 VNSCPHGILG--AKGPIYKYPDAQNECRPCHENCCTQCKGPELQDCLGQAEVLMKPHLV 643
Qy 655 VSAVVGILLVVVLGVVGGILIKRQOKIR-KYTHRRLLQETELVEPLTPSGAMPNQAMR 713
Db 644 IAVTVG--LAVILMILGSGFLYWRGRIQNKRAMRYLERGESIEPLDPS--EKANKVLAR 700
Qy 714 ILKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYV 773
Db 701 IFKETELRKVLGSGVGTGVHKGWVLPAGESIKIPVCIKVIEDKSGRQSFQAVTDHMLA 760
Qy 774 MAGVSPYVSRLLGICLTSTVQLVTLQMPYGCLLDHRVNRGRGLSQDLNMCWQIAKGM 833

Db 761 VGSIDHAHIVRLGLCPSSQLQVLYPLGLSLLDHVKQHRETLPQQLLLNWGVQIAKGM 820
Qy 834 SYLEDVRLVHRDLAARNVLKSPNHNKVTIDFGLARLLDIDETEHADGKGKVPKIMWALES 893
Db 821 YYLESHSVHRDLALRNVLKSPSQVQVADFGVADLLPPDDKQLLHSEAKTPIKWALES 880
Qy 894 ILRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYM 953
Db 881 IHFGKYTHQSDVWSYGVTVWELMTFGAEPYAGRLAEIPDLLEKGERLAQOQICTIDVYM 940
Qy 954 IMVKWMIDSCRFRRELSEFMRMARDPQRFVVIQNEIDGPASPLDSTFYRSLLEDDDD 1013
Db 941 VMVKWMIDENIRTFKELANEFTRMARDPPRYLVIKRAS-GPQTP--PAAEPSVLTKE 997
Qy 1014 MGDLDVAEYLVPOQGFPCPDPAFCAGGMVHRRSSSTRSGGDLTLGLEPSEB----- 1068
Db 998 L-----QEALEPEL-----DLDDLLEABEEGLATS 1023
Qy 1069 -----EAPRSPARSEG-----AGSDVFDGLGMAAKGLOSLPTH 1105
Db 1024 LGSALSUPTGTLTRPGSQSLSPSSGYMPMNQSSLGEACLDLSAVLGGREGQFSRPISLH- 1082
Qy 1106 PSPLQRYSEDPTVLPSETDGVV---APL-----TC-----SPOE---VNOPTDV 1145
Db 1083 PIPGR-----PASSESGHVTCSEALQEKVUCRSRSRSPRGRDSATHSQRS 1135
Qy 1146 RPQPPSPREG-----LPAARPAGATLERAKTLSP-GKNGVV-----KVVFAP 1187
Db 1136 LLTPVTLSPGLEEDGNGVMPDTHLRGASSREGTLSSVGLSSVLGTEEDD----- 1191
Qy 1188 GGAIVENPEYLTPOGGAAPQPHPP 1210
Db 1192 ---EYEVYMNKRKRGSP-PRPP 1209

RESULT 11
TVFVUV
A;Title: protein-tyrosine kinase (EC 2.7.1.112) erbB - avian leukosis virus
N;Contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase
C;Species: avian leukosis virus, ALV
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 11-Jun-1999
C;Accession: B00643; A00643
R;Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1985
A;Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A;Reference number: A00643; MUID:85228222; PMID:2988784
A;Accession: B00643
A;Molecule type: mRNA
A;Residues: 1-698 <Nil>
A;Cross-references: GB:M10066; GB:M13881; NID:G211749; PIDN:AAA48763.1; PID:G211750
A;Note: in Genbank entry CHKERBBF, release 109.0, the source is designated as Gallus gal
C;Comment: This protein is synthesized as a gag-env-erbB protein.
C;Genetics:
A;Gene: gag-env-erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific p
F;1-6/Product: gag protein (fragment) #status predicted <GAG>
F;7-59/Product: env protein (fragment) #status predicted <ENV>
F;60-698/Product: protein-tyrosine kinase erbB #status predicted <ERB>
F;194-459/Domain: protein kinase homology <KIN>
F;202-210/Region: protein kinase ATP-binding motif
F;229/Active site: Lys #status predicted

Query Match 25.9%; Score 1766.5; DB 1; Length 698;
Best Local Similarity 25.2%; Pred. No. 1.2e-65;
Matches 374; Conservative 80; Mismatches 137; Indels 125; Gaps 18;

Qy 578 GPEADQCVACAHYKDPFPCVACRSPGVKPDLSYMPIWKFPDEGACQPCPINCTHSCVDL 637
Db 60 GP--DHCWKCAHFIDGPHCVKACFAGVLGENDTL-VMKYADANAVCQLCHPNCITRGCKGP 116
Qy 638 DDKCPAEQASPLTISVAVV-GILLVVLGVVFGILIKRQKIRKYWRLLQSTEL 696

Db 117 GLEGP---NGSKTPSIAAGVGGGLCLVGVGIGLYLRRR-HIVKRTLRLLRLQREL 172
Qy 697 VEPLTPSGAMPNOAQRILKETELRKVKVGLSGAFGVYKGIWIPDGENVKIPVAIKVL 756
Db 173 VEPLTPSGEAPNOAHLRIKETEFKKVKVGLSGAFGVYKGLWIPGEKVKIPVAIKEL 232
Qy 757 ENTSPKANKILDEAYVMAGVGSPPYVSRLLGICLTSTVQLVQLMPYGCILLDHVRENRR 816
Db 233 EATSPKANKILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCILLDIYREHKDN 292
Qy 817 LGSODLLNWCMOAKGMSYLEDLVRLHRLDAARNVLKSPNHNKVTIDFGLARLLDIDETE 876
Db 293 IGSOYLLNWCVCQAKGNVLEERRLHRLDAARNVLKTPQHKVITDFGLAKLGADEKE 352
Qy 877 YHADGGKVPKIMWALESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLE 936
Db 353 YHAEGGKVPKIMWALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIIPASEISSVLE 412
Qy 937 KGBRLPOPPICITIDVIMVWKCMWIDSECRPRELSEFMRMARDPQRFVVIQ-NEDLG 995
Db 413 KGBRLPOPPICITIDVIMVWKCMWIDADSRPKFRELIAEFSEKWARDPPRYLVITQGD RMH 472
Qy 996 PASPLDSTFYRSLLEDDDDMGDLVAEYLVPOQGFPCPDPAFCAGGMVHRRSSSTRSG 1055
Db 473 LPSTTSKSFRTLWEEDEMDIVDAEYLVPHQGF-----NSPST--- 513
Qy 1056 GGDLTGLSEPESEBAPRSP-----APSEGAGSDVFDGLGMAAKGLOSLPTHDPSPLO 1110
Db 514 -----SRTPLLSLSATSNNATNCID-----RNGQGHVPVREDSFVQ 550
Qy 1111 RYSEDPVPIPSET--DGYVAPLTCSPQEPYVNOPTVVRPQPPRPGPLPAARPAGATLE 1168
Db 551 RYSSDPTGNPLEESIDDDGFL-----PAPEYVNO--LMPKKPS----- 585
Qy 1169 RAKTLPSPKGVVVKDVF-----AFGGAIVENPEYLTPOGGAAPQPHPPAP 1213
Db 586 ----TAMVQHIYNNISLTAISKLPMSRYQNSHSTAVDNPEYL-----NTNQSPLA 633
Qy 1214 SPADFNLXYWDQ-----DPPE-----RGAPPSTFKGTPTAENPEYLGLDVP 1254
Db 634 KTFPESPYWIQSGNHQINLDNFDYQDFLPNETKPNGLLKVPAENPEYLRVAAP 689

RESULT 12
TVYUHV
A;Title: protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain H)
C;Species: avian erythroblastosis virus
C;Date: 18-Apr-1994 #sequence_revision 18-Apr-1994 #text_change 11-Jun-1999
C;Accession: A00644; A38022
R;Yamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.
Cell 35, 71-78, 1983
A;Title: The erbB gene of avian erythroblastosis virus is a member of the src gene fami
A;Reference number: A00644; MUID:84026539; PMID:6313229
A;Accession: A00644
A;Molecule type: DNA
A;Residues: 1-604 <YAW>
A;Cross-references: GB:K01216; NID:G209676; PIDN:AAA42400.1; PID:G209678
R;Debuire, B.; Henry, C.; Benaisa, M.; Biserte, G.; Claverie, J.M.; Saule, S.; Martin,
Science 224, 1456-1459, 1984
A;Title: Sequencing the erba gene of avian erythroblastosis virus reveals a new type of
A;Reference number: A38022; MUID:84223957; PMID:6328658
A;Accession: A38022
A;Molecule type: DNA
A;Residues: 1-28 'W', 30-139, 'F', 141-145, 'V', 147-152 <DEB>
A;Cross-references: GB:K02006
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific
F;130-395/Domain: protein kinase homology <KIN>
F;138-146/Region: protein kinase ATP-binding motif
F;165/Active site: Lys #status predicted

Db 715 QYTAIGPY-----CAASPPRSSKITANLD-----NMFIPIITGAVLVPTIC 755
 Qy VVFGI-LIKRROQKIRKVT--MRLLOTELVEPLTPSGAMPNOAMRILKTELKRVKV 725
 Db 756 ILVVVYICROKQAKKETVKTWALSREDSBPLRFSNIGANICKRIVDAELRGKV 815
 Qy 726 LGSAGFTVYKGIWIPGENYKIPVAIKVLRENTSPKANKSILDEAYVMAGVSPYVRL 785
 Db 816 LGMAGFRVYKGVWPEGENYKIPVAIKELLKSTGAESSEBFLREAYIMASEEHVLLKL 875
 Qy 786 LGICLTSTVOLVTLQMPYGCCLLDHVRNRCGLSQDILLNWCQIAKMSYLEDEVRLVHRD 845
 Db 876 LAVCKSSOMMLITQLMPLGCLLDYVRNRRDKIGSKALLNWTQIAKMSYLEEERLVRD 935
 Qy 846 LAARNVLVK--SPNHVKITDFGLARLLDIDETVHADGGKVPKIKWMALESILRRRTHQ 902
 Db 936 LAARNVLRLAGEDH----DFGLAKLLSSDSNEYKAAGGKMPKWLALCIRNVRTSK 991
 Qy 903 SDVMSYGVVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPCTIDVVMIMVKWMD 962
 Db 992 SDVWAFGVTIWELLTFQORPHENIPAKDIPDLIEVGLKLEQECISLDIYCTLLSCHWLD 1051
 Qy 963 SECPRRELVSERSWARDQRFVVIQNEEDLG--PASPLDSTFYRSLLDD----DMGDL 1017
 Db 1052 AAMRPTFKQLTTFVAFARDPGRYLAIGDKFTLPA-----YTSQDEKDLIRKLAPT 1104
 Qy 1018 VDABEYLVPOQGFPCPOPAPCAGGMVHRHRSSTRSGGDLTLGLEPSEBAP----- 1071
 Db 1105 TDGSEATKPDYIQLPKAALGPS-----HRTDCT-----DEMPKLNRYC 1143
 Qy 1072 RSLPAPSEGAGSDVFDG---DLGMAAGKLOSLPHDPSLQRYSEDPTVPLPSETDGYV 1128
 Db 1144 KDPNSKNSSTGDDERDSSAREVGVGNLR-----LDLPVDEDDYL 1182
 Qy 1129 APLTCSPOPEYVQPDVVRPPSPREGPLPAARFAGATLERAKTLSPGKGVVQDVFAFG 1188
 Db 1183 MP-TCQPGPNNNNNM-----NPNQNNMAAVGVAAGYM-----DLIGVP 1220
 Qy 1189 GAVENPEYL---TPQGAAPQPH-----PPPAFSP-AFDNLXYWD 1224
 Db 1221 VSDNPEYLLNAQTLGVGESPIQTGTGIPVMGGPGTMEVKVPMGPGSEPTSSDHEYND 1279

RESULT 14

S35745

Protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus

C;Species: avian erythroblastosis virus

C;Date: 03-Mar-1994 #sequence_revision 26-May-1995 #text_change 28-Feb-1997

C;Accession: S35745

R;Vennstrom, B.

submitted to the EMBL Data Library, March 1993

A;Reference number: S35743

A;Accession: S35745

A;Molecule type: DNA

A;Residues: 1-544 <VEN>

A;Cross-references: EMBL:X12707

C;Genetics:

A;Gene: erbB

C;Superfamily: epidermal growth factor receptor; protein kinase homology

C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific P

F;135-400/Domain: protein kinase homology <KIN>

F;143-151/Region: protein kinase ATP-binding motif

F;170/Active site: Lys #status predicted

Query Match 24.2%; Score 1647; DB 2; Length 544;

Best Local Similarity 54.9%; Pred. No. 7.3e-61;

Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15;

Qy 578 GPEADQCVACAHYKDPFVCVARGPSGVKPDLSYMPIWKFPPDEGACQPCPINCTHSCVDL 637

Db 1 GP--DHCMKCAHFDIGPHCVKACFAGVLGENDTL-VWKYADANAVCOLCHPNCNTRGCKGP 57

Qy 638 DDGCPAERASPLTSIVSAV--GILLVVVLGVVFGILIKRROOKIRKYTMRLLOTEL 696
 Db 58 GLEGCP---NGSKTPSIAAGWGLLCLVVGVLGIGLYLRRR-HIVKRKTLRLLQREL 113
 Qy 697 VEDLTSGAMPNOAMRILKTELKRVKVLGSGAGFTVYKGIWIPGENYKIPVAIKVL 756
 Db 114 VEDLTSGAMPNOAMRILKTELKRVKVLGSGAGFTVYKGIWIPGENYKIPVAIKVL 173
 Qy 757 ENTSPKANKEILDEAYVMAGVSPYVRLIGICLTSTVOLVTLQMPYGCCLLDHVRNRCR 816
 Db 174 EATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVOLITOLMPYGCCLLDYREHKDN 233
 Qy 817 LGSODLLNWCQIAKMSYLEDEVRLVHRDLAARNVLKSNHVKITDFGLARLLDIDETE 876
 Db 234 IGSQYLLNWCQIAKMSYLEDEVRLVHRDLAARNVLKSNHVKITDFGLARLLDIDETE 293
 Qy 877 YHAGGGKVPKIKWMALESILRRRTHQSDVMSYGVVWELMTFGAKPYDGPAREIPDLLE 936
 Db 294 YHAGGGKVPKIKWMALESILRRRTHQSDVMSYGVVWELMTFGAKPYDGPAREIPDLLE 353
 Qy 937 KGERLPQPPCTIDVVMIMVKWMDSECRPRRELVSERSWARDQRFVVIQ-NEEDLG 995
 Db 354 KGERLPQPPCTIDVVMIMVKWMDSECRPRRELVSERSWARDQRFVVIQ-NEEDLG 413
 Qy 996 PASPLDSTFYRSLLDDDMGDLVDABEYLVPOQGFPCPOPAPCAGGMVHRHRSSTRSG 1055
 Db 414 LPSPTDSKFTYRLMEBEDMEDIVDAEYLVPHQGF-----NSPST--- 454
 Qy 1056 GGDLTIGLEPSEBAPRSL-----APSEGAGSDVFDGDLGMAAGKLOSLPHDPSPLQ 1110
 Db 455 -----SKTPLLSSLSATSNNSATNCDRNGG-----H----- 481
 Qy 1111 RYSEDPTEVLPSETDGYVAPLTCSPQPEYVQPDVVRPPSPREGPLPAARFAGAT-LE 1169
 Db 482 -----PVREDGFL-----PAPEYVQ--LMPKPESTAMVQIYNYISLTAISK 523
 Qy 1170 AKTLSPGKGVVQDVFAFGAVENPEYL 1197
 Db 524 LPIDSRYN-----SHSTAVDNPEYL 544

RESULT 15

S00727

kinase-related transforming protein (erbB) (EC 2.7.1.1) - avian erythroblastosis virus

C;Species: avian erythroblastosis virus

C;Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 28-Feb-1997

C;Accession: S00727

R;Scotting, P.; Vennstrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.

Oncogene Res. 1, 265-278, 1987

A;Title: Common site of mutation in the erbB gene of avian erythroblastosis virus mutant

A;Reference number: S00727; MUID:88217326; PMID:2897102

A;Accession: S00727

A;Molecule type: DNA

A;Residues: 1-545 <SCO>

A;Cross-references: EMBL:X06943

C;Genetics:

A;Gene: erbB

C;Superfamily: epidermal growth factor receptor; protein kinase homology

C;Keywords: ATP; phosphotransferase

F;135-400/Domain: protein kinase homology <KIN>

F;143-151/Region: protein kinase ATP-binding motif

Query Match 24.1%; Score 1640; DB 2; Length 545;

Best Local Similarity 54.9%; Pred. No. 1.4e-60;

Matches 345; Conservative 69; Mismatches 122; Indels 92; Gaps 15;

Qy 578 GPEADQCVACAHYKDPFVCVARGPSGVKPDLSYMPIWKFPPDEGACQPCPINCTHSCVDL 637

Db 1 GP--DHCMKCAHFDIGPHCVKACFAGVLGENDTL-VWKYADANAVCOLCHPNCNTRGCKGP 57

Qy 638 DDGCPAERASPLTSIVSAV--GILLVVVLGVVFGILIKRROOKIRKYTMRLLOTEL 696

Db 58 GLEGCP---NGSKTPSIAAGWGLLCLVVGVLGIGLYLRRR-HIVKRKTLRLLQREL 113

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:11:38 ; Search time 37.9774 Seconds
(without alignments)
403.399 Million cell updates/sec

Title: SEQ4-210-224-12
Perfect score: 6789
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVPV 1255

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 su

```

Database :
A_Geneseq_101002.*
1: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		Length	DB ID	Description
		Match	%			
1	6704	98.7	1255	21	AAE92620	Human heregulin 2
2	6704	98.7	1255	22	AAE12130	Human tyrosine kin
3	6704	98.7	1255	22	AAE60167	HER2 transgene pla
4	6704	98.7	1255	23	AAU74545	Human HER2 (ErB2)
5	6698	98.7	1255	17	AAU01111	HER-2/neu protein.
6	6698	98.7	1255	20	AAW92406	Human HER-2/neu on
7	6698	98.7	1255	21	AAE21198	Human HER-2/neu pr
8	6698	98.7	1255	21	AAV84780	Amino acid sequenc
9	6698	98.7	1255	22	AAE85458	Human HER-2/neu pr
10	6698	98.7	1255	22	AAE88267	HER2/neu amino aci

[illegible]

FT Region /label= Cysteine_rich_domain
FT 210..224
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 250..264
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 324..483
FT /label= Ligand_binding_domain
FT 325..339
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 369..383
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 465..479
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 484..623
FT /label= Cysteine_rich_domain
FT 579..593
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 624..654
FT /label= Transmembrane_domain
FT 632..652
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 653..667
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 655..1010
FT /label= Tyrosine_kinase_domain
FT 661..675
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 695..709
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 710..730
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 1011..1235
FT /label= C-terminal_domain
XX WO200020027-A2.
XX
XX 13-APR-2000.
XX
XX 05-OCT-1999; 99WO-DK00525.
XX
XX 05-OCT-1998; 98DK-0001261.
XX 20-OCT-1998; 98US-0105011.
XX
XX (MEBI-) M & E BIOTECH AS.
XX
XX Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I;
XX Gautam A, Birk P, Karlsson G;
XX WPI; 2000-349917/30.
XX N-PSDB; AAA09455.
XX
XX Inducing immune responses to weakly immunogenic, tumor associated
XX peptide antigens for the treatment of breast and prostate cancer
XX
XX Claim 62; Page 193-198; 220pp; English.
XX
XX This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of
XX Her2 can be used in the Claimed method as an autovaccine to induce a CTL
XX response. Subdominant CTL epitopes, antibody binding regions and
XX cysteine residues involved in disulfide bonds are preserved in the
XX immunogenized forms. Regions suitable for the insertion of foreign T
XX helper epitopes were identified (see features table). The method

CC is used for inducing immune responses against weakly immunogenic
CC cell-associated peptide antigens (PA) such as those associated with
CC cancers (self-proteins), e.g. human prostate specific membrane antigen
CC (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).
CC The method comprises effecting simultaneous presentation by antigen
CC producing cells (APCs) of the animals immune system of: (1) at least 1
CC CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1
CC B-cell group derived from the cell-associated PA; and (2) at least 1
CC first T-helper cell group which is foreign to the animal. Analogues of
CC human PSM, human Her2 and human/murine FGF8b comprising a substantial
CC part of all known and predicted CTL and B-cell epitopes of the respective
CC PA and including at least one foreign T helper epitope are also claimed.
CC The method is used to treat prostate, prostate/breast or breast cancer
CC when the PA is human PSM, FGF8b and Her2, respectively.
XX
SQ Sequence 1255 AA;

Query Match 98.7%; Score 6704; DB 21; Length 1255;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 1242; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLLALLPPGAASTOVCTGTDMLRLPASPTHLDMLRLHYQGCVQGNL 60
DB 1 MELAALCRWGLLLALLPPGAASTOVCTGTDMLRLPASPTHLDMLRLHYQGCVQGNL 60
QY 61 ELTYLPTNASLSFLQDIQEVQGVLIHAHQVROVPLQRLRIVRGTOLEFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSFLQDIQEVQGVLIHAHQVROVPLQRLRIVRGTOLEFEDNYALAVLDNG 120
QY 121 DPLNNTTPVTGASPGGLRELQSLTEILKGGVLIQRNPOLCYQDTILMKDIFHKNNQLA 180
DB 121 DPLNNTTPVTGASPGGLRELQSLTEILKGGVLIQRNPOLCYQDTILMKDIFHKNNQLA 180
QY 181 LTLIDNRSRACHPCSPMCKGSRGWESSEDCQSLRTVTCAGGCARCKGPLPOYIKANSK 240
DB 181 LTLIDNRSRACHPCSPMCKGSRGWESSEDCQSLRTVTCAGGCARCKGPLPOYIKANSK 240
QY 241 FIGITELKHSDDLACLFHNHSGICELHCPALVTYNTDTFESMPNPGRYTFGASCVTACP 300
DB 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVTYNTDTFESMPNPGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSPCARVCYGLGMEHLREVRVTSAN 360
DB 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSPCARVCYGLGMEHLREVRVTSAN 360
QY 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVPETLEEITGYLIYISAWPDSL 420
DB 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVPETLEEITGYLIYISAWPDSL 420
QY 421 DLSVFQNLQVIRGRIILHNGAYSILTLQGLGISWGLSLRELGSGLALIHNTHLCFVHTV 480
DB 421 DLSVFQNLQVIRGRIILHNGAYSILTLQGLGISWGLSLRELGSGLALIHNTHLCFVHTV 480
QY 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQGLCARGHCWGPGTQCVNCSOFLRQEC 540
DB 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQGLCARGHCWGPGTQCVNCSOFLRQEC 540
QY 541 VEECRVLQGLPREYVNAHCLPCHPCQPONGSVTCFGEADQCVACAHYKDPFCVARC 600
DB 541 VEECRVLQGLPREYVNAHCLPCHPCQPONGSVTCFGEADQCVACAHYKDPFCVARC 600
QY 601 PSQVKPDLSPYMPITWPKFPDEGACQPCINCTHSCVDLDDKGCAPAEORASPLTISVAVG 660
DB 601 PSQVKPDLSPYMPITWPKFPDEGACQPCINCTHSCVDLDDKGCAPAEORASPLTISVAVG 660
QY 661 ILLVVLGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
DB 661 ILLVVLGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
QY 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKELDEAYVMAGVGP 780
DB 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKELDEAYVMAGVGP 780

Qy 781 YVSRLLGICLTSTVQLVTLMPYCLLDHVRNRRGLSGQDLNWCQIAKMSYLEDRV 840
Db 781 YVSRLLGICLTSTVQLVTLMPYCLLDHVRNRRGLSGQDLNWCQIAKMSYLEDRV 840
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEYHADGKVPKIMMALESLRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEYHADGKVPKIMMALESLRRFT 900
Qy 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVKCM 960
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVKCM 960
Qy 961 IDSECRPRFRELVSERFARMARDPQFVVIQNEIDLGPASPLDSTFYRSLLEDGMDLVD 1020
Db 961 IDSECRPRFRELVSERFARMARDPQFVVIQNEIDLGPASPLDSTFYRSLLEDGMDLVD 1020
Qy 1021 EYLVPOQGFPCPDPAAGAGMVHRRSSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1080
Db 1021 EYLVPOQGFPCPDPAAGAGMVHRRSSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1080
Qy 1081 AGSDVFDGDLGMAAGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAAGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Qy 1141 NQDVRPQPSPREGPIPAARPAAGATLERAKTILSPGKNGVVKOVFAFGGAVENPEYLTQ 1200
Db 1141 NQDVRPQPSPREGPIPAARPAAGATLERAKTILSPGKNGVVKOVFAFGGAVENPEYLTQ 1200
Qy 1201 GGAAPQHPPPAFSPAFDNLYWDQDPPERGAPSTFKGTPTAENPEYGLDVPV 1255
Db 1201 GGAAPQHPPPAFSPAFDNLYWDQDPPERGAPSTFKGTPTAENPEYGLDVPV 1255

RESULT 2

AAE12130
ID AAE12130 standard; Protein; 1255 AA.
XX
AC AAE12130;
XX
DT 18-DEC-2001 (first entry)
XX
DE Human tyrosine kinase-type receptor, HER-2.
XX
KW Therapeutic compound; major histocompatibility complex; vaccine;
KW antigenic peptide; MHC; immunoregulatory; immune response; HER-2;
KW adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;
KW antigen presenting cell; human; tyrosine kinase-type receptor.
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH Region 774..782
FT /note= "Antigenic epitope"
XX
XX W0200168677-A2.
XX
XX 20-SEP-2001.
XX
XX 16-MAR-2001; 2001WO-US40328.
XX
XX 16-MAR-2000; 2000US-0527487.
PR (GENZ) GENZYME CORP.
XX
XX Nicolette CA;
XX
XX WFI; 2001-616284/71.
DR N-PSDB; AAD19731.
XX
XX Novel synthetic therapeutic compound for inducing immune response and
PT for use in adoptive immunotherapy, has enhanced binding to major
PT histocompatibility molecules and enhanced immunoregulatory properties

XX
PS
XX
CC The invention relates to synthetic therapeutic compounds (antigenic
CC peptides) with enhanced binding to major histocompatibility complex
CC (MHC) molecules and enhanced immunoregulatory properties relative
CC to their natural counterparts. Compounds of the invention are useful
CC for inducing an immune response in a subject and for use in adoptive
CC immunotherapy. They are useful as components of anti-cancer vaccines
CC and to expand immune effector cells that are specific for cancers
CC characterised by expression of the breast cancer antigen, HER-2.
CC Polynucleotides that encode peptides of the invention are useful as
CC hybridisation probes and as primers for the detection of genes of
CC transcripts that are expressed in antigen presenting cells (APCs), to
CC confirm transduction of polynucleotides into host cells. The present
CC sequence is human tyrosine kinase-type receptor, HER-2. Compounds
CC of the invention are designed based on the HER-2 antigenic peptide
CC (774-782).
XX
SQ Sequence 1255 AA;

Query Match 98.7%; Score 6704; DB 22; Length 1255;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 1242; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
Qy 1 MELAALCRWGLLLALLPPGAASQTCTGTDMLRLPASPEHLDMLRHLYQGCVVQGNL 60
Db 1 MELAALCRWGLLLALLPPGAASQTCTGTDMLRLPASPEHLDMLRHLYQGCVVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIEVQGVYLIHAHQVROVPLQRLIRVGTOLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIEVQGVYLIHAHQVROVPLQRLIRVGTOLFEDNYALAVLDNG 120
Qy 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQNPOLCYQDITLWKDIFHKNNOLA 180
Db 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQNPOLCYQDITLWKDIFHKNNOLA 180
Qy 181 LTLIDTNRSHACHPCSPMCKSGRCWGESSEDCQSLTRTVCAGGCARCKGPLPOYIKANSK 240
Db 181 LTLIDTNRSHACHPCSPMCKSGRCWGESSEDCQSLTRTVCAGGCARCKGPLPDCCHCEQC 240
Qy 241 FTGITELKHSIDCLACLFHNSHIGICELHCPALVTYNTDTFESMPNPGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSIDCLACLFHNSHIGICELHCPALVTYNTDTFESMPNPGRYTFGASCVTACP 300
Qy 301 YNVLSTDVGSCTLVCPHNOEYTAEDGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNVLSTDVGSCTLVCPHNOEYTAEDGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360
Qy 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
Db 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420
Qy 421 DLSVFONLQVIRGRILHNGAYSLTLOGLGISWGLRSLRELGSGLALIHNNTHLCFVHTV 480
Db 421 DLSVFONLQVIRGRILHNGAYSLTLOGLGISWGLRSLRELGSGLALIHNNTHLCFVHTV 480
Qy 481 PWDQLFRNPHQALLHTANRPEDECVEGEGACHQLCARGHCWGPGTQCNCVSQFLRGQEC 540
Db 481 PWDQLFRNPHQALLHTANRPEDECVEGEGACHQLCARGHCWGPGTQCNCVSQFLRGQEC 540
Qy 541 VEECRVLQGLPREYVNNARHCLPCHPECPQNGSVTCFPGPEADOCVCAHYKDPFPCVARC 600
Db 541 VEECRVLQGLPREYVNNARHCLPCHPECPQNGSVTCFPGPEADOCVCAHYKDPFPCVARC 600
Qy 601 PSGVKPDLSYMPIWKFPDDEGACQPCPINCTHSCVDLDDKGCFAEQASPLTSIVSAVVG 660
Db 601 PSGVKPDLSYMPIWKFPDDEGACQPCPINCTHSCVDLDDKGCFAEQASPLTSIVSAVVG 660
Qy 661 ILLVVLGVVFGILLIKRROOKIRKYTMRLLOSETVELPELTPSGAMPNQAQRILKETEL 720
Db 661 ILLVVLGVVFGILLIKRROOKIRKYTMRLLOSETVELPELTPSGAMPNQAQRILKETEL 720

Qy 721 RKVKVLSGAGFTYKGIWIPDGENVKIPVAIKVLRNTPSPKANKETILDYAYMAGVGP 780
Db 721 RKVKVLSGAGFTYKGIWIPDGENVKIPVAIKVLRNTPSPKANKETILDYAYMAGVGP 780
Qy 781 YVSRLLGICLTSTVQLVTLQMPYGCCLLDHVRENRLGSGDLLNWCQIAKMSYLEDVR 840
Db 781 YVSRLLGICLTSTVQLVTLQMPYGCCLLDHVRENRLGSGDLLNWCQIAKMSYLEDVR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGGKVPKIKWMALESTLRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGGKVPKIKWMALESTLRRFT 900
Qy 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVWKCM 960
Db 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVWKCM 960
Qy 961 IDSECRPRFRELVSFMRMARDPQFVVIQNEIDLGPASPLDSTFYRSLLEDMDGDLVDA 1020
Db 961 IDSECRPRFRELVSFMRMARDPQFVVIQNEIDLGPASPLDSTFYRSLLEDMDGDLVDA 1020
Qy 1021 EYLVPQGGFCPPAPGAGMWHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
Db 1021 EYLVPQGGFCPPAPGAGMWHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
Qy 1081 AGSDVFDGDLGMAAGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAAGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Qy 1141 NQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKGVVVKDVFAGGAVENPEYLTPO 1200
Db 1141 NQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKGVVVKDVFAGGAVENPEYLTPO 1200
Qy 1201 GGAAPQHPHPPAFSPADNLYYWDQDPPERGAPSTFKGTPTAENPEYLGDLVPEV 1255
Db 1201 GGAAPQHPHPPAFSPADNLYYWDQDPPERGAPSTFKGTPTAENPEYLGDLVPEV 1255

RESULT 3
AAB60167
ID AAB60167 standard; Protein; 1255 AA.
XX
AC AAB60167;
XX
DT 03-APR-2001 (first entry)
XX
DE HER2 transgene plasmid construct encoded protein.
XX
KW Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
KW antibody.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200100244-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-US17229.
XX
PR 25-JUN-1999; 99US-0141316.
PR 16-MAR-2000; 2000US-0189844.
XX
PA (GETH) GENENTECH INC.
XX
PI Erickson S, Schwall R;
XX
DR WPI; 2001-061962/07.
DR N-PSDB; AAF24297.
XX
PT Treating tumors, particularly breast cancers, which overexpress an ErbB
PT receptor and does not respond to an anti-ErbB antibody, comprises
PT conjugating the antibody to a maytansinoid -
XX

PS Example 3; Fig 4; 92pp; English.
XX The present invention provides a method of treating cancer by
XX administering a conjugate of anti-ErbB antibody with a maytansinoid. In
XX particular, the antibody is directed against ErbB2 (also known as HER2
XX and p185neu). The method is particularly useful in the treatment of
XX breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
XX colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.
SQ Sequence 1255 AA;

Query Match 98.7%; Score 6704; DB 22; Length 1255;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 1242; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLLALLPPGAASTQVCTGTDMLKRLPASPTHLDMLRLHYQGCVVQGNL 60
Db 1 MELAALCRWGLLLALLPPGAASTQVCTGTDMLKRLPASPTHLDMLRLHYQGCVVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIOEVQGVYLIHNOVROVPLQRLIRVGTOLFDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIOEVQGVYLIHNOVROVPLQRLIRVGTOLFDNYALAVLDNG 120
Qy 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILMKDIFHKNNQLA 180
Db 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILMKDIFHKNNQLA 180
Qy 181 LTLIDNTRBRACHPCSPMCKSGRCWGESSEDCOSLTRTYCAGGCARCKGPLQYIKANSK 240
Db 181 LTLIDNTRBRACHPCSPMCKSGRCWGESSEDCOSLTRTYCAGGCARCKGPLQYIKANSK 240
Qy 241 FIGITELKSDCLACLFHNSHGICELHCPALVTYNTDTFESMPNPGRYTFGASCVTACP 300
Db 241 AGCTGPKISDCLACLFHNSHGICELHCPALVTYNTDTFESMPNPGRYTFGASCVTACP 300
Qy 301 YNYLSTDVSGCTLVCPHNLQEVTAEDGTQCEKCKPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNYLSTDVSGCTLVCPHNLQEVTAEDGTQCEKCKPCARVCYGLGMEHLREVRVTSAN 360
Qy 361 IQEFACKIKIFGSLAPFESFDGDPASNTAPLOPEQLQVFTELEETGVLIIYSAMDPSLP 420
Db 361 IQEFACKIKIFGSLAPFESFDGDPASNTAPLOPEQLQVFTELEETGVLIIYSAMDPSLP 420
Qy 421 DLSVFONLQVIRGRIILHNGAYSITLQGLGISWGLSLRELGLSGLALIHNNHLCFVHTV 480
Db 421 DLSVFONLQVIRGRIILHNGAYSITLQGLGISWGLSLRELGLSGLALIHNNHLCFVHTV 480
Qy 481 PMDQLFRNPHQALLHTANRPEDECVEGLACHQLCARGHCWGPGTQCVCNCSQFLRGQEC 540
Db 481 PMDQLFRNPHQALLHTANRPEDECVEGLACHQLCARGHCWGPGTQCVCNCSQFLRGQEC 540
Qy 541 VEBCRVLOGLPREYVNRHCLPCHPECQONGSVTCFGEADOCVACAHYKDPFPCVARC 600
Db 541 VEBCRVLOGLPREYVNRHCLPCHPECQONGSVTCFGEADOCVACAHYKDPFPCVARC 600
Qy 601 PSGVKPDLSPYMPIWKPDEEGACQPCINCHTSCVDLDDKGCPEARASPLTSIVSAVVG 660
Db 601 PSGVKPDLSPYMPIWKPDEEGACQPCINCHTSCVDLDDKGCPEARASPLTSIVSAVVG 660
Qy 661 ILVVVVLGVVFGILIKRRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAMRILKETEL 720
Db 661 ILVVVVLGVVFGILIKRRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAMRILKETEL 720
Qy 721 RKVKVLSGAGFTYKGIWIPDGENVKIPVAIKVLRNTPSPKANKETILDYAYMAGVGP 780
Db 721 RKVKVLSGAGFTYKGIWIPDGENVKIPVAIKVLRNTPSPKANKETILDYAYMAGVGP 780
Qy 781 YVSRLLGICLTSTVQLVTLQMPYGCCLLDHVRENRLGSGDLLNWCQIAKMSYLEDVR 840
Db 781 YVSRLLGICLTSTVQLVTLQMPYGCCLLDHVRENRLGSGDLLNWCQIAKMSYLEDVR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGGKVPKIKWMALESTLRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGGKVPKIKWMALESTLRRFT 900

Db -841 LVHRDLAARNVVKSPNHVKTDFGLARLLDIDETEVHADGGKVPKWKMALESILRRRT 900
Qy 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVWCWM 960
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVWCWM 960
Qy 961 IDSECRPRFRELVSERFMRDPPORFVVIQNEIDLGPASPLDSTPYRSLLEDGMDGLVDA 1020
Db 961 IDSECRPRFRELVSERFMRDPPORFVVIQNEIDLGPASPLDSTPYRSLLEDGMDGLVDA 1020
Qy 1021 EYLVPQGGFCPPDPAPGAGWVHHRSSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1080
Db 1021 EYLVPQGGFCPPDPAPGAGWVHHRSSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1080
Qy 1081 AGSDVFDGDLGMAAGKQLSLTHDPSPLQRYSDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAAGKQLSLTHDPSPLQRYSDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Qy 1141 NOPDVRPQPSPREGPIPAARPAATLERAKTILSPGKNGVVKOVFAFGGAVENPEYLTPO 1200
Db 1141 NOPDVRPQPSPREGPIPAARPAATLERAKTILSPGKNGVVKOVFAFGGAVENPEYLTPO 1200
Qy 1201 GGAAPQHPHPPAFSPAFDNLVYNDODPPERGAPPSTFKGPTAENPEYLGLDVVPV 1255
Db 1201 GGAAPQHPHPPAFSPAFDNLVYNDODPPERGAPPSTFKGPTAENPEYLGLDVVPV 1255

RESULT 4
AAU74545
ID AAU74545 standard; Protein; 1255 AA.
XX AAU74545;
DT 23-APR-2002 (first entry)
XX Human HER2 (ErbB2) polypeptide.
XX
KW Human; HER2; ErbB; epidermal growth factor receptor; receptor;
KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
KW glial disorder; astrocytal disorder; hypothalamic disorder;
KW glandular disorder; macrophagal disorder; epithelial disorder;
KW stromal disorder; blastocoealic disorder; inflammatory disorder;
KW angiogenic disorder; immunological disorder.
XX Homo sapiens.
OS
XX
PN US2002001587-A1.
XX
XX 03-JAN-2002.
XX
XX 16-MAR-2001; 2001US-0811123.
XX
XX 16-MAR-2000; 2000US-189844P.
PR 05-OCT-2000; 2000US-238327P.
XX
XX (ERIC/) BRICKSON S.
PA (SCHW/) SCHWALL R.
PA (SLIW/) SLIKOWSKI M.
XX
PI Erickson S, Schwall R, Sliwowski M;
XX
XX WPI; 2002-163686/21.
DR N-PSDB; ABK14058.
XX
XX Treating tumour characterised by overexpression of epidermal growth
PT factor receptor, ErbB or cancer in mammal, comprises administering
PT anti-ErbB antibody-maytansinoid conjugate to the mammal -
XX
XX Example 3; Fig 7; 93pp; English.
PS
XX The invention relates to treating a tumour in a mammal, where the tumour

CC is characterised by the overexpression of an epidermal growth factor
CC receptor (ErbB) and does not respond or responds poorly, to treatment
CC with an anti-ErbB antibody, comprising administering to the mammal an
CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for
CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
CC prostate and bladder, preferably breast cancer. The breast cancer is a
CC metastatic breast cancer or an aggressive form of metastatic breast
CC cancer which overexpresses ErbB2. The method is also useful for treating
CC neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal,
CC epithelial stromal, blastocoealic, inflammatory, angiogenic and
CC immunological disorders. This sequence represents the human HER2 (ErbB2)
CC polypeptide of the invention.
XX
SQ Sequence 1255 AA;
Query Match 98.7%; Score 6704; DB 23; Length 1255;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 1242; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
Qy 1 MELAALCRWGLLALLPPGAASQVCTGTDMLKRLPASPEHDLMLRLHLYQGCVVQGNL 60
Db 1 MELAALCRWGLLALLPPGAASQVCTGTDMLKRLPASPEHDLMLRLHLYQGCVVQGNL 60
Qy 61 ELYLPTNASLSFLQDIQEVQGVYLIHQNQVROVPLQRLIRVGTOLFEDNYALAVLDNG 120
Db 61 ELYLPTNASLSFLQDIQEVQGVYLIHQNQVROVPLQRLIRVGTOLFEDNYALAVLDNG 120
Qy 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQNPOLCYODTILAKDIHKNOLA 180
Db 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQNPOLCYODTILAKDIHKNOLA 180
Qy 181 LTLIDNRSRACHPCSPMCKGSRGCESEDCCSLTRTVCCAGCARCKGKPLQYIKANSK 240
Db 181 LTLIDNRSRACHPCSPMCKGSRGCESEDCCSLTRTVCCAGCARCKGKPLTDCCHCEQC 240
Qy 241 FTIGITELKHSIDCLACLFHNSGICELHCPALVTYNTDTFESMNPREGVYTFGASCVTACP 300
Db 241 AAGCTGKGHSIDCLACLFHNSGICELHCPALVTYNTDTFESMNPREGVYTFGASCVTACP 300
Qy 301 YNYLSTDVSGSCTLVCPHNSQVTAEDGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNYLSTDVSGSCTLVCPHNSQVTAEDGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360
Qy 361 IQEFACGCKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFTELEETIGVLIYISAWPDSL 420
Db 361 IQEFACGCKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFTELEETIGVLIYISAWPDSL 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGSIWGLSLRELGLSGLALIHNTLHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGSIWGLSLRELGLSGLALIHNTLHLCFVHTV 480
Qy 481 PWDQLFRNPHQALLHTANRPEDECVEGLACHQLCARGHCWGPGPTQCVNCSOFLRGQEC 540
Db 481 PWDQLFRNPHQALLHTANRPEDECVEGLACHQLCARGHCWGPGPTQCVNCSOFLRGQEC 540
Qy 541 VEECRVLQGLPREYVNRHCLPCHPBCQONGSVTCFGEADOCVCAHYKDPFFCVARC 600
Db 541 VEECRVLQGLPREYVNRHCLPCHPBCQONGSVTCFGEADOCVCAHYKDPFFCVARC 600
Qy 601 PSGVKPDLSPYIWKFPDEEGACQPCPINCTHSCVDLDDKGCPEAQASPLTSIVSAVVG 660
Db 601 PSGVKPDLSPYIWKFPDEEGACQPCPINCTHSCVDLDDKGCPEAQASPLTSIVSAVVG 660
Qy 661 ILLVVLGVVFGILIKRROOKIRKYTMRLLOSTELVEPLTPSGAMPNOAQMRLKETEL 720
Db 661 ILLVVLGVVFGILIKRROOKIRKYTMRLLOSTELVEPLTPSGAMPNOAQMRLKETEL 720
Qy 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKILDEAYVMAGVGP 780
Db 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKILDEAYVMAGVGP 780
Qy 781 YVSRLLIGICLTSTVQLVTQIMPYGCILLDHVRENRRGLSGQDLINWMCQIAKGSYLEDVR 840

```
Db 781 YVSRLLGICLTSTVQLVTQMPYGCILLDHVRNRRGLSGDQLLNWCQIAKNGSYLEDVR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLADIDETEHADGGKVPKIKWMALESILRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLADIDETEHADGGKVPKIKWMALESILRRFT 900
Qy 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVKWM 960
Db 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVKWM 960
Qy 961 IDSECRPRFRELVSERFARMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDGMDLVD 1020
Db 961 IDSECRPRFRELVSERFARMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDGMDLVD 1020
Qy 1021 EYLVPQGGFCPPAPGAGMWHRRSSSTRSGGDLTLGLEPSEERAPRPLASEG 1080
Db 1021 EYLVPQGGFCPPAPGAGMWHRRSSSTRSGGDLTLGLEPSEERAPRPLASEG 1080
Qy 1081 AGSDVFDGLGMGAAGLQSLPTHDPSPLOKYSDDPTVPLPSETDGVVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGLGMGAAGLQSLPTHDPSPLOKYSDDPTVPLPSETDGVVAPLTCSPQPEYV 1140
Qy 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVYKDVAFPGGAVENPEYLTQ 1200
Db 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVYKDVAFPGGAVENPEYLTQ 1200
Qy 1201 GGAAPQHPHPPAPSPAFDNLVYWDQPPERGAPPSTFKGTPTAENPEYLGIDVPV 1255
Db 1201 GGAAPQHPHPPAPSPAFDNLVYWDQPPERGAPPSTFKGTPTAENPEYLGIDVPV 1255
```

RESULT 5

```
AAW01111 standard; Protein; 1255 AA.
AC AAW01111;
XX
XX 01-JAN-1997 (first entry)
XX
XX HER-2/neu protein.
XX
XX HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
KW breast cancer; ovary cancer; colon cancer; lung cancer;
KW prostate cancer; immunisation; tumour; vaccine; vector.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Domain 676..1255
XX FT /label= Intracellular domain
XX FT /note= "Claimed domain, useful for immunisation"
XX
XX W09630514-A1.
XX
XX 03-OCT-1996.
XX
XX 28-MAR-1996; 96WO-US01689.
XX
XX 31-MAR-1995; 95US-0414417.
XX
XX (UNIW ) UNIV WASHINGTON.
XX
XX Cheever MA, Disis ML;
XX
XX WPI; 1996-455361/45.
XX N-PSDB; AAT40739.
XX
XX DNA encoding HER-2-neu poly:peptide(s) - used for prevention or
XX treatment of malignancies with which the HER-2/neu oncogene is
XX associated
XX
XX Claim 2; Page 56-61; 71pp; English.
```

```
XX Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is
CC the product of the HER-2/neu oncogene (see also AAT40739). The
CC protein is over-expressed in various cancers, including breast,
CC ovarian, colon, lung and prostate. The intracellular domain of the
CC protein can be used to immunise an animal against a malignancy with
CC which the oncogene is associated. The polypeptide can be produced
CC in transformed host cells for use in immunisation. Alternatively,
CC animal cells are transfected in vivo or ex vivo with a viral vector
CC that directs expression of the polypeptide.
XX
XX SQ Sequence 1255 AA;
Query Match 98.7%; Score 6698; DB 17; Length 1255;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 1; Mismatches 14; Indels 0; Gaps 0;
Qy 1 MELAALCRWGLLLALLPPGAASTQVCTGDMKRLRLPASPTHLDMLRHLYQGQVQGNL 60
Db 1 MELAALCRWGLLLALLPPGAASTQVCTGDMKRLRLPASPTHLDMLRHLYQGQVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIOEVQGVLIHAHQVROVPLQRLRIVRGTLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIOEVQGVLIHAHQVROVPLQRLRIVRGTLFEDNYALAVLDNG 120
Qy 121 DPLNNTTPTVGTASPGGLRELQRLSLTEILKGGVLIQRPOLQCYQDTILWKDIFHKNNQLA 180
Db 121 DPLNNTTPTVGTASPGGLRELQRLSLTEILKGGVLIQRPOLQCYQDTILWKDIFHKNNQLA 180
Qy 181 LTLIDTNRSRACHPCSPMCKGSRGCESEDQSLTRTVGAGCARCKGPLPOYIKANSK 240
Db 181 LTLIDTNRSRACHPCSPMCKGSRGCESEDQSLTRTVGAGCARCKGPLPDDCCEQC 240
Qy 241 FTGITELKHSDDCLACLFHNSGICELHCPALTYNTDTFESMNPGRVTFEGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALTYNTDTFESMNPGRVTFEGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPHLNHOEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNYLSTDVGSCTLVCPHLNHOEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360
Qy 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFETLEETIGLYISAWPDSL 420
Db 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFETLEETIGLYISAWPDSL 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWLGSLRLSRELGLALIHHTHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWLGSLRLSRELGLALIHHTHLCFVHTV 480
Qy 481 PWDOLFNRPHQALLHTANRPEBECVGEGLACHQLCARGHCWGPGTQCVCNCSQFLRGQEC 540
Db 481 PWDOLFNRPHQALLHTANRPEBECVGEGLACHQLCARGHCWGPGTQCVCNCSQFLRGQEC 540
Qy 541 VEECRVLQGLPREYVNAHCLPCHPECQFNGSVTCFGEADQCVACAHYKDPFPCVARC 600
Db 541 VEECRVLQGLPREYVNAHCLPCHPECQFNGSVTCFGEADQCVACAHYKDPFPCVARC 600
Qy 601 PSGVKPDLISYMPIWKFPDEGACQPCPINCTHSCVDLDDKGCFAEORASPLTSISAVWG 660
Db 601 PSGVKPDLISYMPIWKFPDEGACQPCPINCTHSCVDLDDKGCFAEORASPLTSISAVWG 660
Qy 661 ILLVVVLGVVFGVILIKRROOKIRKYTMRRLLQBELVEPLTPSGAMPNQAQMRILKETEL 720
Db 661 ILLVVVLGVVFGVILIKRROOKIRKYTMRRLLQBELVEPLTPSGAMPNQAQMRILKETEL 720
Qy 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIIPVAIKVLRNTSPKANKSILDEAYVMAGVSP 780
Db 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIIPVAIKVLRNTSPKANKSILDEAYVMAGVSP 780
Qy 781 YVSRLLGICLTSTVQLVTQMPYGCILLDHVRNRRGLSGDQLLNWCQIAKNGSYLEDVR 840
Db 781 YVSRLLGICLTSTVQLVTQMPYGCILLDHVRNRRGLSGDQLLNWCQIAKNGSYLEDVR 840
```

Db 1021 EYLVPQGFCDPAPGAGMWHRRSSSTRSGGDLTLGLEPSEEAAPRPLAPSEG 1080
Qy 1081 AGSDVFDGDLGMAKGLQSLPTHDPSPLOQYSEDPTVPLPSETDGVVAPLTCSPOPEYV 1140
Db 1081 AGSDVFDGDLGMAKGLQSLPTHDPSPLOQYSEDPTVPLPSETDGVVAPLTCSPOPEYV 1140
Qy 1141 NQDVRPQPPSPREGPLPAARPAATLERAATLSPGKNGVVKOVFAFGGAVENPEYLTPO 1200
Db 1141 NQDVRPQPPSPREGPLPAARPAATLERAATLSPGKNGVVKOVFAFGGAVENPEYLTPO 1200
Qy 1201 GGAAPQHPPAPSPAFDNLYWDQPPERGAPPSTPKGTPTAENPYLGLDVPV 1255
Db 1201 GGAAPQHPPAPSPAFDNLYWDQPPERGAPPSTPKGTPTAENPYLGLDVPV 1255

RESULT 9
AAB85458
ID AAB85458 standard; Protein; 1255 AA.
AC AAB85458;
XX
DT 25-SEP-2001 (first entry)
XX Human HER-2/neu protein.
DE Antigen-presenting cell; immunogenic; immune response; HER-2/neu;
DE oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.
KW Homo sapiens.
XX
XX WO200153463-A2.
XX
XX 26-JUL-2001.
XX
XX 19-JAN-2001; 2001WO-US01850.
XX
XX 21-JAN-2000; 2000US-0177545.
XX
XX (CORI-) CORIXA CORP.
XX
XX Cheever MA, Hand-Zimmermann S;
XX
XX WPI; 2001-476112/51.
XX
XX N-PSDB; AAH23392.
XX
XX New antigen-presenting cells, useful as vaccines for eliciting or
XX enhancing an immune response to HER-2/neu protein, particularly useful
XX for treating or preventing cancer, e.g. breast cancer -
XX
XX Claim 2; Page 41-46; 49pp; English.
XX
XX The invention provides an isolated antigen-presenting cell, which
XX expresses at least an immunogenic portion of a polypeptide that produces
XX an immune response to HER-2/neu protein. The antigen-presenting cells are
XX useful as vaccines for eliciting or enhancing an immune response to
XX HER-2/neu protein, particularly in treating or preventing malignancies in
XX which the HER-2/neu oncogene is associated. Specifically, these are
XX useful for treating or preventing cancer, e.g. breast cancer, ovarian,
XX colon, lung or prostate cancers. The present sequence represents
XX the human HER-2/neu protein (also known as p185 or c-erbB2).
XX
XX Sequence 1255 AA;
XX

Query Match. 98.7%; Score 6698; DB 22; Length 1255;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 1; Mismatches 14; Indels 0; Gaps 0;
Qy 1 MELAAICRWGLLLALLPPGAASVCTGCTDMKRLPASPETHLDMRLHYGCGVQGNL 60
Db 1 MELAAICRWGLLLALLPPGAASVCTGCTDMKRLPASPETHLDMRLHYGCGVQGNL 60
Qy 61 ELTYLPTNASLFLQDIOEVQGVLIHAHQVRQVPLQRLIRVGTQLPFDNYALAVLDNG 1200

Db 61 ELTYLPTNASLFLQDIOEVQGVLIHAHQVRQVPLQRLIRVGTQLPFDNYALAVLDNG 120
Qy 121 DPLNNTPTVGTASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNNOLA 180
Db 121 DPLNNTPTVGTASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNNOLA 180
Qy 181 LTLIDTNRGRACHPCSPMCKSGRCWGESSEDCQSLTRTVTCAGGCARCKGPLPOYIKANSK 240
Db 181 LTLIDTNRGRACHPCSPMCKSGRCWGESSEDCQSLTRTVTCAGGCARCKGPLPDDCCEQC 240
Qy 241 FTGITELKHSDDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTACP 300
Db 241 AAGCTGPKFISDDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPHNOEVTAEQGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNYLSTDVGSCTLVCPHNOEVTAEQGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360
Qy 361 IQEFAGCKKI:FGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAMPDLSL 420
Db 361 IQEFAGCKKI:FGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAMPDLSL 420
Qy 421 DLSVFONLQVIRGRILHNGAYSITLQGLGISWGLSLRELGSGLALIIHNTHLCFVHTV 480
Db 421 DLSVFONLQVIRGRILHNGAYSITLQGLGISWGLSLRELGSGLALIIHNTHLCFVHTV 480
Qy 481 PDQLFRNHQALHTANRPEDECVGEGLAGHOLCARGHCWGPGTQCVNCSOFLRGOEC 540
Db 481 PDQLFRNHQALHTANRPEDECVGEGLAGHOLCARGHCWGPGTQCVNCSOFLRGOEC 540
Qy 541 VEECRVLQGLPREYVVARHCLPCHPECQPNQSGVTCFGEADQCVACAHYKDPFVCVARC 600
Db 541 VEECRVLQGLPREYVVARHCLPCHPECQPNQSGVTCFGEADQCVACAHYKDPFVCVARC 600
Qy 601 PSGVKPDLSTYMPIWKFPDEBEGACQPCINCTHSCVDLDDKGCAPORASPLTSIVSAVVG 660
Db 601 PSGVKPDLSTYMPIWKFPDEBEGACQPCINCTHSCVDLDDKGCAPORASPLTSIVSAVVG 660
Qy 661 ILLVVLGVVFGILIKRQOKIRKYTMRLLOTELVEPLTPSGAMPNOAMRILKETEL 720
Db 661 ILLVVLGVVFGILIKRQOKIRKYTMRLLOTELVEPLTPSGAMPNOAMRILKETEL 720
Qy 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKSILDEAYVMAGVGP 780
Db 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKSILDEAYVMAGVGP 780
Qy 781 YVSRLLIGICLTSTVQLVTQMLPYGCLLDHVRENRGLSGODLLNWCQIAKGSYLEDDVR 840
Db 781 YVSRLLIGICLTSTVQLVTQMLPYGCLLDHVRENRGLSGODLLNWCQIAKGSYLEDDVR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRRFT 900
Qy 901 HQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPICTIDVYIMVKCWM 960
Db 901 HQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPICTIDVYIMVKCWM 960
Qy 961 IDSECRPRFELVSEFSRMARDQRFVIVONEDLGPASPLDSTFYRSLLEDDMDGLVDA 1020
Db 961 IDSECRPRFELVSEFSRMARDQRFVIVONEDLGPASPLDSTFYRSLLEDDMDGLVDA 1020
Qy 1021 EYLVPOQGFCDPAPGAGMWHRRSSSTRSGGDLTLGLEPSEEAAPRPLAPSEG 1080
Db 1021 EYLVPOQGFCDPAPGAGMWHRRSSSTRSGGDLTLGLEPSEEAAPRPLAPSEG 1080
Qy 1081 AGSDVFDGDLGMAKGLQSLPTHDPSPLOQYSEDPTVPLPSETDGVVAPLTCSPOPEYV 1140
Db 1081 AGSDVFDGDLGMAKGLQSLPTHDPSPLOQYSEDPTVPLPSETDGVVAPLTCSPOPEYV 1140
Qy 1141 NQDVRPQPPSPREGPLPAARPAATLERAATLSPGKNGVVKOVFAFGGAVENPEYLTPO 1200

Db 961 IDSECRPRFRELSEFMRMDPQRFVVIQNEGLPASPLDSTFYRSLLEDMDGLVDA 1020
QY 1021 EYLVPQGGFFCPDPAPGAGGWHHRSSSTRSGGDLTLGLEPSEEEAPRGLAPSEG 1080
Db 1021 EYLVPQGGFFCPDPAPGAGGWHHRSSSTRSGGDLTLGLEPSEEEAPRGLAPSEG 1080
QY 1081 AGSDVFDGLGMAAAGLQSLPHTDPSLQRYSDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGLGMAAAGLQSLPHTDPSLQRYSDPTVPLPSETDGYVAPLTCSPQPEYV 1140
QY 1141 NQDVRPQPSREGPLPAAPAGATLERAKTILSPGKNGVVKOVFAFGGAVENPEYLTPO 1200
Db 1141 NQDVRPQPSREGPLPAAPAGATLERAKTILSPGKNGVVKOVFAFGGAVENPEYLTPO 1200
QY 1201 GGAAPQHPPPAFSPAFDNLVYWDQDPPERGAPSTFKGTPTAENPEYLGLDVVP 1255
Db 1201 GGAAPQHPPPAFSPAFDNLVYWDQDPPERGAPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 8

AAAY84780

ID AAAY84780 standard; Protein; 1255 AA.

AC AAAY84780;

DT 08-AUG-2000 (first entry)

XX Amino acid sequence of the SPLICE erbB-2 receptor protein.

XX SPLICE erbB-2 receptor protein; cell transformation disorder; cancer;
XX tumor cell proliferation; tissue degeneration; arthropathy;
XX bone resorption; inflammatory disease; degenerative disorder;
XX wound healing.

OS Homo sapiens.

XX WP1; 2000-303768/26.

XX N-PSDB; AAA14812.

XX 13-APR-2000.

XX 01-OCT-1999; 99WO-CA00912.

XX 02-OCT-1998; 98US-0165192.

XX (UYMC-) UNIV MCMASTER.

XX Muller WJ, Siegel PM;

XX WP1; 2000-303768/26.

XX N-PSDB; AAA14812.

XX Nucleic acid encoding an erbB 2 receptor protein designated SPLICE
XX erbB-2, inhibitors of the protein are useful for treatment of cancer -
XX Claim 3; Fig 2; 60pp; English.

XX The present sequence represents a SPLICE erbB-2 receptor protein. The
XX protein has an in-frame deletion of 16 amino acids, 2 of which are
XX conserved cysteine residues, compared to the unspliced protein. The
XX erbB-2 polynucleotide is used to construct probes for detecting
XX disorders of cell transformation such as cancer. Antibodies to the
XX protein may be used to detect SPLICE erbB-2 in a sample. Agents
XX (e.g. antisense oligonucleotides) which inhibit the expression of
XX SPLICE erbB-2 are useful for reducing tumor cell proliferation and
XX treating cancer. Substances which stimulate SPLICE erbB-2 are useful
XX for treating conditions involving damaged cells including conditions
XX in which degeneration of tissue occurs, such as arthropathy, bone
XX resorption, inflammatory diseases, degenerative disorders of the
XX central nervous system and wound healing.

XX Sequence 1255 AA;

SQ

Query Match 98.7%; Score 6698; DB 21; Length 1255;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLALPPGAASTQVCTGDMKRLPASPTHLDMLRHLYQGCVVQGNL 60
Db 1 MELAALCRWGLLALPPGAASTQVCTGDMKRLPASPTHLDMLRHLYQGCVVQGNL 60
QY 61 ELYTLPTNASLSFLQDIQEVQGYVLIAHNVQROVPLQRLIRVGTOLFEDNYALAVLDNG 120
Db 61 ELYTLPTNASLSFLQDIQEVQGYVLIAHNVQROVPLQRLIRVGTOLFEDNYALAVLDNG 120
QY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILMKDIFHKKNQLA 180
Db 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILMKDIFHKKNQLA 180
QY 181 LTLIDNTRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVCAAGCARCKGPLPQYIKANSK 240
Db 181 LTLIDNTRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVCAAGCARCKGPLPQYIKANSK 240
QY 241 FIGITELKHSDDLCLAHFNHSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLAHFNHSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQCEKSKPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQCEKSKPCARVCYGLGMEHLREVRVTSAN 360
QY 361 IQEFAGCKKIFGSLAPLPESFDGDPASNTAPLOPEQLQVFETLEETIGLYISANWDSLP 420
Db 361 IQEFAGCKKIFGSLAPLPESFDGDPASNTAPLOPEQLQVFETLEETIGLYISANWDSLP 420
QY 421 DLSVFONLOVIRGRILHNGAYSILTLQGLIGISWGLSRLRELGLGLAIHHNTHLCFVHTV 480
Db 421 DLSVFONLOVIRGRILHNGAYSILTLQGLIGISWGLSRLRELGLGLAIHHNTHLCFVHTV 480
QY 481 PWDQLFRNPQHALLHTANRPEDECVGEGLAHQLCARGHCWGFGPTQCVCNCSQFLRGQEC 540
Db 481 PWDQLFRNPQHALLHTANRPEDECVGEGLAHQLCARGHCWGFGPTQCVCNCSQFLRGQEC 540
QY 541 VEECRVLQGLPREYVNAHCLFCHPECCQNGSVTCFGEADQCACAHYKDPFPFCVARC 600
Db 541 VEECRVLQGLPREYVNAHCLFCHPECCQNGSVTCFGEADQCACAHYKDPFPFCVARC 600
QY 601 PSGVKPDLSPYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCAPAEORASPLTSIYSAVVG 660
Db 601 PSGVKPDLSPYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCAPAEORASPLTSIYSAVVG 660
QY 661 ILLVVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNOAQRILKETEL 720
Db 661 ILLVVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNOAQRILKETEL 720
QY 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVRENTSPKANKEILDEAYVMAGVGSF 780
Db 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVRENTSPKANKEILDEAYVMAGVGSF 780
QY 781 VYSRLGLICTSTVOLVTOLMPYGLLDHVRNRLGSDQLLNMCQIAGKMSYLEDVYR 840
Db 781 VYSRLGLICTSTVOLVTOLMPYGLLDHVRNRLGSDQLLNMCQIAGKMSYLEDVYR 840
QY 841 LVHRDLAARNVLKSPNPKITDFGLARLLDDIDETEHADGKVPKWMALESIILRRRT 900
Db 841 LVHRDLAARNVLKSPNPKITDFGLARLLDDIDETEHADGKVPKWMALESIILRRRT 900
QY 901 HQSDVMSYGVTVWELMTGAKPYDGI PAREIPDLLEKGERLPQPICTIDVTVMYVKWM 960
Db 901 HQSDVMSYGVTVWELMTGAKPYDGI PAREIPDLLEKGERLPQPICTIDVTVMYVKWM 960
QY 961 IDSECRPRFRELVSFSSRMARDPQRFVVIQNEGLPASPLDSTFYRSLLEDMDGLVDA 1020
Db 961 IDSECRPRFRELVSFSSRMARDPQRFVVIQNEGLPASPLDSTFYRSLLEDMDGLVDA 1020
QY 1021 EYLVPQGGFFCPDPAPGAGGWHHRSSSTRSGGDLTLGLEPSEEEAPRGLAPSEG 1080

Qy 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVWKWM 960
Db 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVWKWM 960
Qy 961 IDSECRPRFELVSEFRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDMGLDVA 1020
Db 961 IDSECRPRFELVSEFRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDMGLDVA 1020
Qy 1021 EYLVPQGFPCPDPAAGAGWVHRRSSSTRSGGDLTLGLEPSEEEAPRSPAPSEG 1080
Db 1021 EYLVPQGFPCPDPAAGAGWVHRRSSSTRSGGDLTLGLEPSEEEAPRSPAPSEG 1080
Qy 1081 AGSDVFDGLGMGAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVAPLTCSPOPEYV 1140
Db 1081 AGSDVFDGLGMGAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVAPLTCSPOPEYV 1140
Qy 1141 NOPDVRQPPSPREGPLPAARPAGATLERAKTILSPGKGVVVDVAFGGAVENPEYLTPO 1200
Db 1141 NOPDVRQPPSPREGPLPAARPAGATLERAKTILSPGKGVVVDVAFGGAVENPEYLTPO 1200
Qy 1201 GGAAPQHPHPPAFSPAFDNLYYWDQDPPERGAPSTFKGTPTAENPEYLGLDVVPV 1255
Db 1201 GGAAPQHPHPPAFSPAFDNLYYWDQDPPERGAPSTFKGTPTAENPEYLGLDVVPV 1255

RESULT 7

AAB21198
ID AAB21198 standard; protein; 1255 AA.

XX AC AAB21198;
XX DT 12-JAN-2001 (first entry)

XX DE Human HER-2/neu protein.
XX KW Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;
XX KW breast cancer; prostate cancer; ovarian cancer; lung cancer;
XX KW colon cancer.

XX OS Homo sapiens.

XX PN WO200044899-A1.

XX PD 03-AUG-2000.

XX PF 28-JAN-2000; 2000WO-US02164.

XX PR 29-JAN-1999; 99US-0117976.

XX PA (CORI-) CORIXA CORP.

XX PA (SMIK) SMITHKLINE BEECHAM.

XX PI Cheever MA, Gheysen D;

XX DR WPI; 2000-505976/45.

XX DR N-PSDB; AAA89736.

XX PT HER-2/neu extracellular domain/phosphorylation domain fusion proteins
XX PT useful for vaccinating against breast, ovarian, colon, lung and
XX PT prostate cancers -
XX PS Claim 52; Fig 7; 128pp; English.
XX CC The present sequence is the human HER-2/neu protein. It is a member of
XX CC the tyrosine kinase family of receptor-like glycoproteins and shows
XX CC homology to the epidermal growth factor receptor (EGFR). It probably
XX CC plays a part in cell growth and/or differentiation. The HER-2/neu
XX CC gene is an oncogene. An HER-2/neu fusion protein comprising a
XX CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
XX CC domain may be used to treat or prevent cancer by eliciting or
XX CC enhancing an immune response to the HER-2/neu protein. It may be used
XX CC to treat malignancies such as breast, ovarian, colon, lung and
XX CC prostate cancers, and may be used as an antigen to vaccinate against

CC these neoplasias.
XX SQ Sequence 1255 AA;
Query Match 98.7%; Score 6698; DB 21; Length 1255;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 1; Mismatches 14; Indels 0; Gaps 0;
Qy 1 MELAALCRGGLLALLPPGAASQTCTGTDMLRLPASPEHLDMLRLHYQGCVVQGNL 60
Db 1 MELAALCRGGLLALLPPGAASQTCTGTDMLRLPASPEHLDMLRLHYQGCVVQGNL 60
Qy 61 ELTYLPTNASLSEFLQDIQEVQGVYVLIHQNQVQVPLQRLIRVRGTOLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSEFLQDIQEVQGVYVLIHQNQVQVPLQRLIRVRGTOLFEDNYALAVLDNG 120
Qy 121 DPLNNTPTVTGASPGGLRELOLRSLTEILKGGVLIQORNPOLCYODTILKNDIIFHKNNOLA 180
Db 121 DPLNNTPTVTGASPGGLRELOLRSLTEILKGGVLIQORNPOLCYODTILKNDIIFHKNNOLA 180
Qy 181 LTLIDTNRGRACHPCSPMCKSGSCWSESSEDQSLTRTVTCAGGCARCKGPLPOYIKANSK 240
Db 181 LTLIDTNRGRACHPCSPMCKSGSCWSESSEDQSLTRTVTCAGGCARCKGPLPTDCCHCEQC 240
Qy 241 FIGITELKHSDDLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRAVTSAN 360
Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRAVTSAN 360
Qy 361 IQBFACKKIFGSLAFLPESFDGDPASNTAPLOEQVFEETLEETITGVLYISAWPDSL 420
Db 361 IQBFACKKIFGSLAFLPESFDGDPASNTAPLOEQVFEETLEETITGVLYISAWPDSL 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISMGLSLRELGLSLALIHNNTHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISMGLSLRELGLSLALIHNNTHLCFVHTV 480
Qy 481 PWDOLFERNPHCALLHTANRPEDECVGEGLAACHOLCARGHCWGPGTQCVCNCSOFLRQEC 540
Db 481 PWDOLFERNPHCALLHTANRPEDECVGEGLAACHOLCARGHCWGPGTQCVCNCSOFLRQEC 540
Qy 541 VEECRVLQGLFREYVNAHCLPCHPECQONGSVTCFGEADOCVACAHYKDPFVCVARC 600
Db 541 VEECRVLQGLFREYVNAHCLPCHPECQONGSVTCFGEADOCVACAHYKDPFVCVARC 600
Qy 601 PSGVKPDLSTNPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPEAQRASPLTSIVSAVVG 660
Db 601 PSGVKPDLSTNPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPEAQRASPLTSIVSAVVG 660
Qy 661 ILLVVLGVVFGILIKRQOKIRKYTWRLLOLQTELVEPLTPSGAMPNQOMILKETEL 720
Db 661 ILLVVLGVVFGILIKRQOKIRKYTWRLLOLQTELVEPLTPSGAMPNQOMILKETEL 720
Qy 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMAGVSP 780
Db 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMAGVSP 780
Qy 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRGLSGODLLNWCWQIAKMSYLEDVR 840
Db 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRGLSGODLLNWCWQIAKMSYLEDVR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKVMMALESILRRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKVMMALESILRRRFT 900
Qy 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVWKWM 960
Db 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVWKWM 960
Qy 961 IDSECRPRFELVSEFRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDMGLDVA 1020

QY 341 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWKMALESILRRFT 900
 DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWKMALESILRRFT 900
 QY 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPDPPICTIDVYMWKWM 960
 DB 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPDPPICTIDVYMWKWM 960
 QY 961 IDSECRPRFRELSEFSEFARMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDGMDLVD 1020
 DB 961 IDSECRPRFRELSEFSEFARMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDGMDLVD 1020
 QY 1021 EYLVPOQGFCDPAPAGAGWVHRHRSSTRSGGDLTLGLEPSEEEAPRPLAPSEG 1080
 DB 1021 EYLVPOQGFCDPAPAGAGWVHRHRSSTRSGGDLTLGLEPSEEEAPRPLAPSEG 1080
 QY 1081 AGSDVFDGDLGMGAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAELTCSPOPEYV 1140
 DB 1081 AGSDVFDGDLGMGAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAELTCSPOPEYV 1140
 QY 1141 NQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAFGGAVENPEYLTPO 1200
 DB 1141 NQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAFGGAVENPEYLTPO 1200
 QY 1201 GGAAPQHPHPPAPSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVPV 1255
 DB 1201 GGAAPQHPHPPAPSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVPV 1255

RESULT 6

AAW92406
 ID AAW92406 standard; Protein; 1255 AA.
 AC AAW92406;
 DT 21-APR-1999 (first entry)
 XX Human HER-2/neu oncogene protein.
 XX HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;
 KW malignancy; treatment; tumour.
 XX Homo sapiens.
 OS
 FH Key
 FT Region
 FT 676..1255
 FT /note= "region which elicits immune response"
 XX
 PN US5869445-A.
 XX
 XX 09-FEB-1999.
 XX
 XX 01-APR-1996; 96US-0625101.
 XX
 PR 01-APR-1996; 96US-0625101.
 PR 17-MAR-1993; 93US-0033644.
 PR 12-AUG-1993; 93US-0106112.
 PR 31-MAR-1995; 95US-0414417.
 XX
 XX (UNIW) UNIV WASHINGTON.
 XX
 XX Cheever MA, Disis ML;
 XX WPI; 1999-152835/13.
 DR N-PSDB; AAX01912.
 XX

XX Use of HER-2/neu polypeptides - for eliciting an immune response to
 PT an HER-2/neu associated malignancy, particularly for treating or
 PT preventing tumours

XX Claim 3; Column 31-38; 26pp; English.

XX This sequence represents the human HER-2/neu oncogene protein. A fragment

CC of this protein is used in a method for eliciting or enhancing an immune
 CC response to HER-2/neu protein. The polypeptide can stimulate T cells and
 CC B cells to produce an immune response to the HER-2/neu protein. The
 CC method can be used for immunisation against a malignancy in which the
 CC HER-2/neu oncogene is associated and in the treatment of an existing
 CC tumour, or to prevent tumour occurrence or reoccurrence.
 XX

SQ Sequence 1255 AA;

Query Match 98.7%; Score 6698; DB 20; Length 1255;
 Best Local Similarity 98.8%; Pred. No. 0;
 Matches 1240; Conservative 1; Mismatches 14; Indels 0; Gaps 0;
 QY 1 MELAALCRWGLLLALLPPGAASQVCTGTDMLRLPASPETHDMLRLHLYQGCVVQGNL 60
 DB 1 MELAALCRWGLLLALLPPGAASQVCTGTDMLRLPASPETHDMLRLHLYQGCVVQGNL 60
 QY 61 ELYLPTNASLFLQDIQEVQGVLIHNOVQVPLQRLIRVRGTOLFEDNYALAVLDNG 120
 DB 61 ELYLPTNASLFLQDIQEVQGVLIHNOVQVPLQRLIRVRGTOLFEDNYALAVLDNG 120
 QY 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRPOLCYODTILWKDIFHKKNOLA 180
 DB 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRPOLCYODTILWKDIFHKKNOLA 180
 QY 181 LTLIDTNRGRACHPCSPCKGSRGWESSEDQSLTRTVCAGGCARCKGKPLPOYIKANSK 240
 DB 181 LTLIDTNRGRACHPCSPCKGSRGWESSEDQSLTRTVCAGGCARCKGKPLPOYIKANSK 240
 QY 241 FTITELKHSDDLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 DB 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 QY 301 YNYLSTDVCSCTLVCPHNOEVTAEQGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360
 DB 301 YNYLSTDVCSCTLVCPHNOEVTAEQGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360
 QY 361 IOEFAGCKKIFGSLAFPLPESFDGDPASNTAPLOPEOLQVFETLEETIGVLYISAWPDSL 420
 DB 361 IOEFAGCKKIFGSLAFPLPESFDGDPASNTAPLOPEOLQVFETLEETIGVLYISAWPDSL 420
 QY 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWGLRSLRELGLSGLALIHNNHLCFVHTV 480
 DB 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWGLRSLRELGLSGLALIHNNHLCFVHTV 480
 QY 481 PWDQLFRNPHQALLHTANRPEDECYVGEGLACHQLCARGHCWGPPTQVCNCSQFLRGQEC 540
 DB 481 PWDQLFRNPHQALLHTANRPEDECYVGEGLACHQLCARGHCWGPPTQVCNCSQFLRGQEC 540
 QY 541 VEECRVLOGLPREYVNAHCLPCHPECOQONGSVTCFGEADQCVACAHYKDPFPCVARG 600
 DB 541 VEECRVLOGLPREYVNAHCLPCHPECOQONGSVTCFGEADQCVACAHYKDPFPCVARG 600
 QY 601 PSGVKPDLSPYMPIWKFPDSEBACQPCPNCTHSCVDLDDKGCPEAQASPLTSIVSAVVG 660
 DB 601 PSGVKPDLSPYMPIWKFPDSEBACQPCPNCTHSCVDLDDKGCPEAQASPLTSIVSAVVG 660
 QY 661 ILLVVVLGVVFGILIKRQOKIRKYMRLLOETELVEPLTPSGAMPNQAQRILKETEL 720
 DB 661 ILLVVVLGVVFGILIKRQOKIRKYMRLLOETELVEPLTPSGAMPNQAQRILKETEL 720
 QY 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVALKVLRENTSPKANKEILDEAYVMAVGSP 780
 DB 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVALKVLRENTSPKANKEILDEAYVMAVGSP 780
 QY 781 YVSRLLGICLTSTVOLVTQLMPYGLLDHVRNRLGSLQDILLNMCWQIAKMSYLEDYR 840
 DB 781 YVSRLLGICLTSTVOLVTQLMPYGLLDHVRNRLGSLQDILLNMCWQIAKMSYLEDYR 840
 QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWKMALESILRRFT 900
 DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWKMALESILRRFT 900

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:05:49 ; Search time 0.755229 Seconds
(without alignments)
1909.378 Million cell updates/sec

Title: US-09-806-703A-12

Perfect score: 74
Sequence: 1 QYIKANSKFIGITEL 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 73:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	74	100.0	1315	1 BTCLTN	tentoxilysin (EC 3
2	44.5	60.1	244	2 S29982	class II histocomp
3	44	59.5	66	2 S31029	gene 84 protein -
4	43	58.1	180	2 G86826	diamine N-acetyltr
5	43	58.1	899	2 T42976	hypothetical prote
6	42.5	57.4	1060	2 S06286	major merozoite su
7	42.5	57.4	1086	2 S16752	major merozoite su
8	42.5	57.4	1701	2 A54498	major merozoite su
9	42.5	57.4	1701	2 A26868	major merozoite su
10	42.5	57.4	1726	1 SAZQGM	major merozoite su
11	42.5	57.4	1726	2 A45948	major merozoite su
12	42	56.8	1333	2 S38635	blastoplia polyprot
13	41	55.4	123	2 G48677	Ig heavy chain V-D
14	41	55.4	447	2 H97146	siderophore/Surfac
15	41	55.4	899	2 C36812	hypothetical prote
16	40.5	54.7	245	2 S29980	class II histocomp
17	40	54.1	79	2 D85794	hypothetical prote
18	40	54.1	194	2 G64026	lacyl-carrier-prot
19	40	54.1	601	1 A55485	oligopeptidase (EC
20	40	54.1	601	2 G86840	oligopeptidase
21	40	54.1	644	2 S46746	hypothetical prote
22	39	52.7	102	2 PH1491	Ig heavy chain V r
23	39	52.7	119	2 PH1516	Ig heavy chain V r
24	39	52.7	119	2 PH1518	Ig heavy chain V r
25	39	52.7	119	2 PH1519	Ig heavy chain V r
26	39	52.7	123	2 F48677	Ig heavy chain V-D
27	39	52.7	135	2 PH1494	Ig heavy chain V r
28	39	52.7	140	2 PH1488	Ig heavy chain V r
29	39	52.7	189	2 G97978	conserved hypoten

adenylate kinase (type II restrictio probable acetyl-Co type IIS restricti type IIS restricti hypothetical prote hypothetical prote hypothetical prote hypothetical prote tRNA-pseudouridine probable hexosyltr hypothetical prote hypothetical prote endo-1,4-beta gluc cellulase (EC 3.2. methyl-accepting c hemagglutinin-neur

30 39 52.7 213 1 KIYWC

31 39 52.7 326 2 B71808

32 39 52.7 349 2 T43043

33 39 52.7 423 2 F64690

34 39 52.7 505 2 C90569

35 38 51.4 188 2 H71875

36 38 51.4 188 2 A64639

37 38 51.4 256 2 F64472

38 38 51.4 287 2 F70361

39 38 51.4 381 2 F71196

40 38 51.4 383 2 T51466

41 38 51.4 424 2 T29127

42 38 51.4 501 2 A86158

43 38 51.4 501 2 T52135

44 38 51.4 561 2 E82395

45 38 51.4 572 1 HNNZ80

ALIGNMENTS

RESULT 1

BTCLTN

tentoxilysin (EC 3.4.24.68) precursor - Clostridium tetani

N;Alternate names: tetanus neurotoxin

C;Species: Clostridium tetani

C;Date: 31-Mar-1988 #sequence revision 31-Mar-1988 #text change 03-Jun-2002

C;Accession: A25689;A25757; A25194; B25194; A60759; S69348; S09364

R;Eisel, U.; Jarausch, W.; Goretzki, K.; Henschen, A.; Engels, J.; Weller, U.; Hudel, M., EMBO J. 5, 2495-2502, 1986

A;Title: Tetanus toxin: primary structure, expression in E. coli, and homology with botu

A;Reference number: A25689; MUID:87053814; PMID:3536478

A;Accession: A25689

A;Molecule type: DNA

A;Residues: 1-1315 <EIS>

A;Cross-references: GB:X04436; NID:940769; PIDN:CAA28033.1; PID:940770

R;Fairweather, N.F.; Lyness, V.A.

Nucleic Acids Res. 14, 7809-7812, 1986

A;Title: The complete nucleotide sequence of tetanus toxin.

A;Reference number: A25757; MUID:87040747; PMID:3774547

A;Accession: A25757

A;Molecule type: DNA

A;Residues: 1-1315 <FAI>

A;Cross-references: GB:X06214; NID:940773; PIDN:CAA29564.1; PID:940774

A;Experimental source: strain CN3911

R;Fairweather, N.F.; Lyness, V.A.; Pickard, D.J.; Allen, G.; Thomson, R.O.

J. Bacteriol. 165, 21-27, 1986

A;Title: Cloning, nucleotide sequencing, and expression of tetanus toxin fragment C in E

A;Reference number: A25194; MUID:86085672; PMID:3510187

A;Accession: A25194

A;Molecule type: DNA

A;Residues: 743-1315 <FA2>

A;Cross-references: GB:M12739; NID:9144920; PIDN:AAA23282.1; PID:9144921

A;Accession: B25194

A;Molecule type: protein

A;Residues: 865-894 <FA3>

R;Matsuda, M.; Lei, D.L.; Sugimoto, N.; Ozutsumi, K.; Okabe, T.

Infect. Immun. 57, 3588-3593, 1989

A;Title: Isolation, purification, and characterization of fragment B, the NH-2-terminal t

A;Reference number: A60759; MUID:90035436; PMID:2478476

A;Accession: A60759

A;Molecule type: protein

A;Residues: 461-475 <MAT>

R;Demotz, S.; Lanzavecchia, L.; Eisel, U.; Niemann, H.; Widmann, C.; Corradin, G.

J. Immunol. 142, 394-402, 1989

A;Title: Delineation of several DR-restricted tetanus toxin T cell epitopes.

A;Reference number: JS0098; MUID:89093918; PMID:2463305

A;Contents: annotation; epitope region

R;Schlavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta, B.R.

Nature 359, 832-835, 1992

A;Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolytic

A;Reference number: S27125; MUID:93063293; PMID:1331807

A;Contents: annotation

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1992

C:Genetics: 4
A:Gene: 84
A:Start codon: GTG

Query Match 59.5%; Score 44; DB 2; Length 66;
Best Local Similarity 57.1%; Pred. No. 0.58;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 YIKANSKFIGITEL 15
DB 50 YIKRNGKFGVTWEV 63
||| ||| ||| |

RESULT 4
G86826
diamine N-acetyltransferase (EC 2.3.1.57) [imported] - Lactococcus lactis subsp. lactis
N:Alternate names: spermidine acetyltransferase
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 17-May-2002
C:Accession: G86826
R:R:Botolin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarre, K.; Weissenbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: G86826
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-180 <STO>
A:Cross-references: GB:AE005176; PID:gl2724622; PIDN:AAK05713.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: Ygff
C:Superfamily: Escherichia coli ribosomal-protein-alanine N-acetyltransferase rimJ
C:Keywords: acyltransferase; coenzyme A

Query Match 58.1%; Score 43; DB 2; Length 180;
Best Local Similarity 69.2%; Pred. No. 2.5;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 IKANSKFIGITEL 15
DB 65 IEANDTFIGIVEL 77
||| ||| ||| |||

RESULT 5
T42976
hypothetical protein 63 - ateline herpesvirus 3 (strain 73)
C:Species: ateline herpesvirus 3
A:Variety: strain 73
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
C:Accession: T42976
R:R:Albrecht, J.C.; Fleckenstein, B.
submitted to the EMBL Data Library, August 1998
A:Description: Primary structure of the herpesvirus ateles genome.
A:Reference number: Z22274
A:Accession: T42976
A>Status: preliminary; translated from GB/EMBL/DDBU
A:Molecule type: DNA
A:Residues: 1-899 <ALB>
A:Cross-references: EMBL:AF083424; PIDN:AAC95587.1
A:Experimental source: strain 73

Query Match 58.1%; Score 43; DB 2; Length 899;
Best Local Similarity 64.3%; Pred. No. 13;
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 QYIKANSKFIGITE 14
DB 124 QYITNSFTTGOTE 137
||| ||| ||| |||

RESULT 6

C;Keywords: surface antigen

Query Match 57.4%; Score 42.5; DB 2; Length 1701;
Best Local Similarity 60.0%; Pred. No. 30;
Matches 9; Conservative 1; Mismatches 1; Gaps 1;

Qy 1 QYIKANSKFI-GITE 14
|:::| | | | |
Db 1001 QFVKSNSKVITGLTE 1015

RESULT 9
A:Accession: A26868
C:Species: Plasmodium falciparum
C:Date: 19-Nov-1998 #sequence_revision 19-Nov-1988 #text_change 09-Jun-2000
E:Tanabe, K.; Mackay, M.; Goman, M.; Scaife, J.G.
J. Mol. Biol. 195, 273-287, 1987
A:Title: Allelic dimorphism in a surface antigen gene of the malaria parasite Plasmodium falciparum
A:Reference number: A26868; MUID:88011243; PMID:3079521
A:Accession: A26868
A:Molecule type: DNA
A:Residues: 1-1701 <TAN>
C:Superfamily: major merozoite surface antigen
C:Keywords: surface antigen
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1701/Product: major merozoite surface antigen #status predicted <MAT>

Query Match 57.4%; Score 42.5; DB 2; Length 1701;
Best Local Similarity 60.0%; Pred. No. 30;
Matches 9; Conservative 1; Mismatches 1; Gaps 1;

Qy 1 QYIKANSKFI-GITE 14
|:::| | | | |
Db 1001 QFVKSNSKVITGLTE 1015

RESULT 10
SAZOOM
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (strain
N:Alternate names: 195K glycoprotein
C:Species: Plasmodium falciparum
C:Date: 30-Sep-1987 #sequence_revision 31-Mar-1991 #text_change 09-Jun-2000
C:Accession: A23386; S06361
R:Weber, J.L.; Leininger, W.M.; Lyon, J.A.
Nucleic Acids Res. 14, 3311-3323, 1986
A:Title: Variation in the gene encoding a major merozoite surface antigen of the human malarial parasite Plasmodium falciparum
A:Reference number: A23386; MUID:86205236; PMID:3517809
A:Accession: A23386
A:Molecule type: DNA
A:Residues: 1-1104 <WEB1>
A:Cross-references: EMBL:X03831
R:Weber, J.L.; Sim, B.K.L.; Lyon, J.A.; Wolff, R.
Nucleic Acids Res. 16, 1206, 1988
A:Title: Merozoite surface protein sequence from the Camp strain of the human malaria parasite Plasmodium falciparum
A:Reference number: S06361; MUID:88143999; PMID:3278296
A:Accession: S06361
A:Molecule type: DNA
A:Residues: 1104-1726 <WEB2>
A:Cross-references: EMBL:X03831
C:Comment: The merozoite stages of different strains have strain-specific surface antigens. The merozoite stages of Plasmodium falciparum has three stages: sporozoite, merozoite, and gametocyte. The merozoite stage is the most abundant in the blood of the infected host.
C:Superfamily: major merozoite surface antigen
C:Keywords: glycoprotein; merozoite; surface antigen; tandem repeat
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1726/Product: major merozoite surface antigen #status predicted <MAT>
F:67-87, 91-96, 100-105, 109-120/Region: 3-residue repeats (S-G-T)
F:757-765/Region: 3-residue repeats (T-E-E)
F:133, 272, 501, 567, 638, 827, 839, 944, 990, 1016, 1114, 1221, 1613, 1658/Binding site: carbohydrate binding site

Query Match 57.4%; Score 42.5; DB 1; Length 1726;
Best Local Similarity 60.0%; Pred. No. 31;

Oy 1 QYKANSKFIGITE 14
Db 124 QYITSNATFTGLSE 137

Search completed: July 22, 2003, 08:12:43
Job time : 2.75523 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:05:49 ; Search time 1.51046 Seconds
(without alignments)

2046.206 Million cell updates/sec

Title: US-09-806-703A-12

Perfect score: 74

Sequence: 1 QYIKANSKFIGITEL 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 21:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organellae:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	74	100.0	1310	2 Q93N27	Q93n27 clostridium
2	44.5	60.1	60	7 Q31585	Q31585 salmo salar
3	44.5	60.1	71	7 Q9XRJ9	Q9xrj9 salvelinus
4	44.5	60.1	85	7 Q951S2	Q951s2 salmo salar
5	44.5	60.1	85	7 Q95HY1	Q95hy1 salmo salar
6	44.5	60.1	86	7 Q95HX4	Q95hx4 salmo salar
7	44.5	60.1	244	7 Q31590	Q31590 salmo salar
8	44	59.5	545	10 Q9XG37	Q9xg37 guillardia
9	43.5	58.8	67	7 Q31578	Q31578 salmo salar
10	43	58.1	180	16 Q9CF66	Q9cf66 lactococcus
11	43	58.1	250	9 Q9MCL7	Q9mcl7 streptococ
12	43	58.1	252	9 Q9XJ8	Q9xj8 lactococcus
13	43	58.1	302	11 Q9CRV4	Q9crv4 mus musculu
14	43	58.1	309	11 Q9CYD2	Q9cyd2 mus musculu
15	43	58.1	899	12 Q9YTK4	Q9ytk4 ateline her
16	42.5	57.4	1087	5 Q25961	Q25961 plasmodium

17	42.5	57.4	1694	5 Q9NHX1	Q9nhx1 plasmodium
18	42.5	57.4	1694	5 Q9TZT5	Q9tzt5 plasmodium
19	42.5	57.4	1704	5 Q9TZT4	Q9tzt4 plasmodium
20	42.5	57.4	1720	5 Q25922	Q25922 plasmodium
21	42	56.8	358	17 Q8TZU8	Q8tzu8 pyrococcus
22	42	56.8	1333	5 Q24262	Q24262 drosophila
23	41.5	56.1	84	13 Q9DEK4	Q9dek4 coregonus s
24	41.5	56.1	85	7 Q95IS3	Q95is3 salmo salar
25	41.5	56.1	85	7 Q95IR2	Q95ir2 salmo salar
26	41.5	56.1	149	7 Q31495	Q31495 oncorhynch
27	41.5	56.1	216	7 Q9GJH0	Q9gh0 salmo trutt
28	41.5	56.1	216	7 Q9GJG9	Q9gj9 salmo trutt
29	41	55.4	187	16 Q8R9Q7	Q8r9q7 thermoanaer
30	41	55.4	247	11 Q9D3B9	Q9d3b9 mus musculu
31	41	55.4	384	4 Q9HD07	Q9hd07 homo sapien
32	41	55.4	384	16 Q8RCF9	Q8rcf9 thermoanaer
33	41	55.4	447	16 Q97HK7	Q97hk7 clostridium
34	41	55.4	532	5 Q96671	Q96671 drosophila
35	41	55.4	540	5 Q9VU53	Q9vu53 drosophila
36	40.5	54.7	67	7 Q31577	Q31577 salmo salar
37	40.5	54.7	67	7 Q31581	Q31581 salmo salar
38	40.5	54.7	67	7 Q31582	Q31582 salmo salar
39	40.5	54.7	71	7 Q9XRH1	Q9xrh1 salvelinus
40	40.5	54.7	71	7 Q9XRH6	Q9xrh6 salvelinus
41	40.5	54.7	84	13 Q9DEJ6	Q9dej6 coregonus s
42	40.5	54.7	85	7 Q95IS9	Q95is9 salmo salar
43	40.5	54.7	85	7 Q95IS8	Q95is8 salmo salar
44	40.5	54.7	85	7 Q95IS6	Q95is6 salmo salar
45	40.5	54.7	85	7 Q95IS5	Q95is5 salmo salar

ALIGNMENTS

RESULT 1

Q93N27 Q93N27 PRELIMINARY; PRT; 1310 AA.
 AC Q93N27;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE Tetanus toxin (Fragment).
 OS Clostridium tetani.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
 OC Clostridiales; Clostridiaceae; Clostridium.
 OX NCBI_TaxID=1513;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Shuman Z., Dianliang L.;
 RT "Cloning and sequence analysis of tetanus toxin gene."
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF389424; AAK72964.2; -;
 DR InterPro: IPR000395; Bontoxilysin.
 DR InterPro: IPR001064; Crystallin.
 DR InterPro: IPR000130; Zn_Mtpeptdee.
 DR Pfam: PF01742; Peptidase M27; 1.
 DR ProDom: PD001963; Bontoxilysin; 1.
 DR PROSITE: PS00225; CRYSTALLIN BETA GAMMA; UNKNOWN_1.
 DR PROSITE: PS00142; ZINC_PROTEASE; UNKNOWN_1.
 FT NON_TER 1 1
 FT NON_TER 1310 1310
 SQ SEQUENCE 1310 AA; 150316 MW; 9EADDC914418E450 CRC64;

Query Match 100.0%; Score 74; DB 2; Length 1310;
 Best Local Similarity 100.0%; Pred. No. 0.00019;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QYIKANSKFIGITEL 15

Db 831 QYIKANSKFIGITEL 845

RESULT 2

```
Q31585
ID Q31585 PRELIMINARY; PRT; 60 AA.
AC Q31585;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE (DB02) MHC class II beta 1 (Fragment).
OS Salmo salar (Atlantic salmon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
OX NCBI_TaxID=8030;
RN [1]
RP SEQUENCE FROM N.A.
RA Grimholt U., Oleaker I., de Vries Lindstrom C., Lie O.;
RT "A study of polymorphism in the MHC class II beta 1 and MHC class I
RT alpha 2 domain exons of Atlantic salmon (Salmo salar).";
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; L24953; AAA49597.1; -.
DR InterPro; IPR000353; MHC_II_beta.
DR Pfam; PF00969; MHC_II_beta.1.
DR ProDom; PD000328; MHC_II_beta.1.
KW Glycoprotein; MHC II; Transmembrane.
FT NON_TER 1
FT NON_TER 60
SQ SEQUENCE 60 AA; 6776 MW; 5D4F3449060940E2 CRC64;

Query Match 60.1%; Score 44.5; DB 7; Length 60;
Best Local Similarity 55.6%; Pred. No. 1.6;
Matches 10; Conservative 3; Mismatches 2; Indels 3; Gaps 1;

QY 1 QYIKANS---KFIGITEL 15
Db 16 EYIRFNSTVGKFGVGYTEL 33

RESULT 3
Q9XRJ9 PRELIMINARY; PRT; 71 AA.
AC Q9XRJ9;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MHC class II beta 1 (Fragment).
GN SANA.
OS Salvelinus namaycush (Lake trout).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salvelinus.
OX NCBI_TaxID=8040;
RN [1]
RP SEQUENCE FROM N.A.
RA Dorschner M.O., Duris T., Phillips R.B.;
RT "Diversity of a Lake trout Mhc class II Gene.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF130026; AAD20899.1; -.
DR InterPro; IPR000353; MHC_II_beta.
DR Pfam; PF00969; MHC_II_beta.1.
DR ProDom; PD000328; MHC_II_beta.1.
KW Glycoprotein; MHC II; Transmembrane.
FT NON_TER 1
FT NON_TER 71
SQ SEQUENCE 71 AA; 8002 MW; E3095286582A9F2D CRC64;

Query Match 60.1%; Score 44.5; DB 7; Length 71;
Best Local Similarity 55.6%; Pred. No. 1.9;
Matches 10; Conservative 3; Mismatches 2; Indels 3; Gaps 1;

QY 1 QYIKANS---KFIGITEL 15
Db 14 EYIRFNSTVGKFGVGYTEL 31

Q31585
ID Q31585 PRELIMINARY; PRT; 85 AA.
AC Q31585;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE MHC class II beta chain (Fragment).
OS Salmo salar (Atlantic salmon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
OX NCBI_TaxID=8030;
RN [1]
RP SEQUENCE FROM N.A.
RA Landry C., Bernatchez L.;
RT "Comparative analysis of population structure across environments and
RT geographic scales at Major Histocompatibility Complex and
RT microsatellite in Atlantic salmon (Salmo salar).";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF373699; AAK61882.1; -.
DR InterPro; IPR000353; MHC_II_beta.
DR Pfam; PF00969; MHC_II_beta.1.
DR ProDom; PD000328; MHC_II_beta.1.
KW Glycoprotein; MHC; MHC II; Transmembrane.
FT NON_TER 1
FT NON_TER 85
SQ SEQUENCE 85 AA; 9743 MW; 3214E01AD1B66AC5 CRC64;

Query Match 60.1%; Score 44.5; DB 7; Length 85;
Best Local Similarity 55.6%; Pred. No. 2.2;
Matches 10; Conservative 3; Mismatches 2; Indels 3; Gaps 1;

QY 1 QYIKANS---KFIGITEL 15
Db 33 EYIRFNSTVGKFGVGYTEL 50

RESULT 5
Q9SHY1 PRELIMINARY; PRT; 85 AA.
AC Q9SHY1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE MHC class II B antigen (Fragment).
GN DBI.
OS Salmo salar (Atlantic salmon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
OX NCBI_TaxID=8030;
RN [1]
RP SEQUENCE FROM N.A.
RA Langefors A., Lohm J., von Schantz T.;
RT "Allelic polymorphism in MHC class II B in four populations of
RT Atlantic salmon (Salmo salar).";
RL Immunogenetics 53:329-336(2001).
DR EMBL; AF104370; AAL04002.1; -.
DR InterPro; IPR000353; MHC_II_beta.
DR Pfam; PF00969; MHC_II_beta.1.
DR ProDom; PD000328; MHC_II_beta.1.
KW Glycoprotein; MHC; MHC II; Transmembrane.
FT NON_TER 1
FT NON_TER 85
SQ SEQUENCE 85 AA; 9723 MW; 27C9F7931F1F01C4 CRC64;

Query Match 60.1%; Score 44.5; DB 7; Length 85;
Best Local Similarity 55.6%; Pred. No. 2.2;
Matches 10; Conservative 3; Mismatches 2; Indels 3; Gaps 1;

QY 1 QYIKANS---KFIGITEL 15
```

```
Db          :||: || ||: ||||
33 EYIRFNSTVGKFGVGYTEL 50

RESULT 6
Q95HX4      PRELIMINARY;      PRT;      86 AA.
AC Q95HX4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE MHC class II B antigen (Fragment).
GN DBI.
OS Salmo salar (Atlantic salmon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
OX NCBI_TaxID=8030;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21383619; PubMed=11491536;
RA Langefors A., Lohm J., von Schantz T.;
RT "Allelic polymorphism in MHC class II B in four populations of
RT Atlantic salmon (Salmo salar)";
RL Immunogenetics 53:329-336(2001).
DR EMBL; AF104377; AAL04009.1; -.
DR InterPro; IPR000353; MHC_II_beta.
DR Pfam; PF00969; MHC_II_beta; 1.
DR ProDom; PD000328; MHC_II_beta; 1.
KW Glycoprotein; MHC; MHC II; Transmembrane.
FT NON_TER
FT TER
FT TER
SQ SEQUENCE 86 AA; 9912 MW; E5097729F681F149 CRC64;

Query Match      60.1%; Score 44.5; DB 7; Length 86;
Best Local Similarity 55.6%; Pred. No. 2.2;
Matches 10; Conservative 3; Mismatches 2; Indels 3; Gaps 1;

QY 1 QYIKANS---KFIGITEL 15
: ||: || ||: ||||
33 EYIRFNSTVGKFGVGYTEL 50

Db          :||: || ||: ||||
33 EYIRFNSTVGKFGVGYTEL 50

RESULT 7
Q31590      PRELIMINARY;      PRT;      244 AA.
AC Q31590;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MHC class II.
GN MHC-SASA CLASS II B.
OS Salmo salar (Atlantic salmon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
OX NCBI_TaxID=8030;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=LEUKOCYTE;
RX MEDLINE=93170890; PubMed=8436418;
RA Hordvik I., Grimholt U., Fosse V.M., Lie Y., Endresen C.;
RT "Cloning and sequence analysis of cDNAs encoding the MHC class II a-
RT chain in Atlantic salmon, Salmo salar.";
RL Immunogenetics 37:437-441(1993).
DR EMBL; X70166; CAA49725.1; -.
DR HSSP; P01888; IBMG.
DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 1.
DR Pfam; PF00969; MHC_II_beta; 1.
DR ProDom; PD000328; MHC_II_beta; 1.
```

```
DR SMART; SM00407; IGcl1; 1.
KW Glycoprotein; MHC II; Transmembrane.
SQ SEQUENCE 244 AA; 27449 MW; 496CB9EA9D73765C CRC64;

Query Match      60.1%; Score 44.5; DB 7; Length 244;
Best Local Similarity 55.6%; Pred. No. 6.4;
Matches 10; Conservative 3; Mismatches 2; Indels 3; Gaps 1;

QY 1 QYIKANS---KFIGITEL 15
: ||: || ||: ||||
51 EYIRFNSTVGKFGVGYTEL 68

Db          :||: || ||: ||||
51 EYIRFNSTVGKFGVGYTEL 68

RESULT 8
Q9XG37      PRELIMINARY;      PRT;      546 AA.
AC Q9XG37;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Hypothetical 66.2 kDa protein.
OS Guillardia theta (Cryptomonas phi).
OC Eukaryota; Cryptophyta; Cryptomonadales; Guillardia.
OX NCBI_TaxID=55529;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20087226; PubMed=10618395;
RA Zauner S., Fraunholz M., Wastl J., Penny S.L., Beaton M.,
RA Cavalier-Smith T., Maier U., Douglas S.;
RT "Chloroplast protein and centrosomal genes, a tRNA intron, and odd
RT telomeres in an unusually compact eukaryotic genome, the cryptomonad
RT nucleomorph.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:200-205(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21233349; PubMed=11323671;
RA Douglas S., Zauner S., Fraunholz M., Beaton M., Penny S., Deng L.,
RA Wu X., Reith M., Cavalier-Smith T., Maier U.;
RT "The highly reduced genome of an enslaved algal nucleus.";
RL Nature 410:1091-1096(2001).
DR EMBL; AJ010592; CAB40403.1; -.
KW Hypothetical protein.
SQ SEQUENCE 546 AA; 66218 MW; 7303950F632BE6F2 CRC64;

Query Match      59.5%; Score 44; DB 10; Length 546;
Best Local Similarity 50.0%; Pred. No. 18;
Matches 7; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY 2 YIKANSKFIGITEL 15
: ||: || ||: ||||
445 FIKNSRFRWRLTEI 458

Db          :||: || ||: ||||
445 FIKNSRFRWRLTEI 458

RESULT 9
Q31578      PRELIMINARY;      PRT;      67 AA.
AC Q31578;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE (DB03) MHC class II beta 1 (Fragment).
OS Salmo salar (Atlantic salmon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
OX NCBI_TaxID=8030;
RN [1]
RP SEQUENCE FROM N.A.
RA Grimholt U., Olsaker I., de Vries Lindstrom C., Lie O.;
RT "A study of polymorphism in the MHC class II beta 1 and MHC class I
RT alpha 2 domain exons of Atlantic salmon (Salmo salar).";
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; L24929; AAA49550.1; -.
```

```

DR InterPro; IPR000353; MHC II beta.
DR Pfam; PF00969; MHC II beta; 1.
DR ProDom; PD000328; MHC II beta; 1.
KW Glycoprotein; MHC II; Transmembrane.
FT NON_TER 67
FT 1
SQ SEQUENCE 67 AA; 7449 MW; 42771AEDBABA6626 CRC64;

Query Match 58.1%; Score 43.5; DB 7; Length 67;
Best Local Similarity 50.0%; Pred. No. 2.6; Mismatches 4; Indels 3; Gaps 1;
Matches 9; Conservative 4;

QY 1 QYIKANS---KFIGITEL 15
Db 16 EYVENSTVGKFGYTEL 33

RESULT 10
Q9CF66 PRELIMINARY; PRT; 180 AA.
AC Q9CF66;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Spermidine acetyltransferase (EC 2.3.1.57).
GN YOFF OR L11615.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=1137471;
RA Bolotin A., Winkler P., Mauger S., Jaillon O., Malarne K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403."
RL Genome Res. 11:731-753 (2001).
DR ENBL; AE006391; AAK05713.1; -.
DR InterPro; IPR000182; Gcn5acetyltransf.
DR Pfam; PF00583; Acetyltransf; 1.
DR Transferase; Complete proteome.
KW SEQUENCE 180 AA; 21022 MW; 6DBD148524C0DF3C CRC64;

Query Match 58.1%; Score 43; DB 16; Length 180;
Best Local Similarity 69.2%; Pred. No. 8.8;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 IKANSKFIGITEL 15
Db 65 IEANDTFIGIVEL 77

RESULT 11
Q9MCL7 PRELIMINARY; PRT; 250 AA.
AC Q9MCL7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE ORF13.
GN ORF13.
OS Streptococcus thermophilus bacteriophage 7201.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OX NCBI_TaxID=112023;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20088830; PubMed=10620678;
RA Stanley E., Walsh L., van der Zwet A., Fitzgerald G.F.,
RA van Sinderen D.;
RT "Identification of four loci isolated from two Streptococcus
RT thermophilus phage genomes responsible for mediating bacteriophage
resistance.";
RT PEMS Microbiol. Lett. 182:271-277 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Stanley E., Walsh L., Fitzgerald G.F., van Sinderen D.;
RA Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR ENBL; AF145054; AAF43506.1; -.
DR InterPro; IPR001091; CNA Metransf.
DR InterPro; IPR002295; D21N6_mtfase.
DR Pfam; PF01555; N6_N4_Mtase; 1.
DR PRINTS; PR00506; D21N6MTFRASE.
DR PRINTS; PR00508; S21N4MTFRASE.
SQ SEQUENCE 250 AA; 28744 MW; 725EA3B2B56B0D7F CRC64;

Query Match 58.1%; Score 43; DB 9; Length 250;
Best Local Similarity 57.1%; Pred. No. 12;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 QYIKANSKFIGITE 14
Db 131 QVLKANKIVGATE 144

RESULT 12
Q9XJEB PRELIMINARY; PRT; 252 AA.
AC Q9XJEB;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative methylase.
OS Lactococcus lactis bacteriophage Tuc2009.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OX NCBI_TaxID=35241;
RN [1]
RP SEQUENCE FROM N.A.
RA van Sinderen D., van de Guchte M., Seegers J.F.M.L., Fitzgerald G.F.;
RA "Molecular analysis of the temperate lactococcal phage Tuc2009."
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR ENBL; AF109874; AAD37103.1; -.
DR InterPro; IPR001091; CNA Metransf.
DR InterPro; IPR002295; D21N6_mtfase.
DR Pfam; PF01555; N6_N4_Mtase; 1.
DR PRINTS; PR00506; D21N6MTFRASE.
DR PRINTS; PR00508; S21N4MTFRASE.
KW Methyltransferase.
SQ SEQUENCE 252 AA; 29357 MW; 4D5A59FF47363948 CRC64;

Query Match 58.1%; Score 43; DB 9; Length 252;
Best Local Similarity 57.1%; Pred. No. 12;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 QYIKANSKFIGITE 14
Db 130 QVLKANKIVGATE 143

RESULT 13
Q9CRV4 PRELIMINARY; PRT; 302 AA.
AC Q9CRV4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE 3110030A04Rik protein (Fragment).
GN 3110030A04RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:05:43 ; Search time 0.395043 Seconds
(without alignments)
1574.881 Million cell updates/sec

Title: US-09-806-703A-12

Perfect score: 74

Sequence: 1 QYKANSKFIGITEL 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	74	100.0	1314	1	TETX CLOTE
2	44	59.5	66	1	VG8A_BPML5
3	42.5	57.4	1682	1	MSPI_PLAFL3
4	42.5	57.4	1701	1	MSPI_PLAFL3
5	42.5	57.4	1701	1	MSPI_PLAFL3
6	42.5	57.4	1726	1	MSPI_PLAFL3
7	42.5	57.4	1726	1	MSPI_PLAFL3
8	41	55.4	204	1	PYRC_SERMA
9	41	55.4	384	1	LE12_THETN
10	41	55.4	899	1	V120_HSVSA
11	40	54.1	194	1	ACPD_HABIN
12	40	54.1	601	1	PEFL_LACLC
13	40	54.1	601	1	PEPF_LACLC
14	40	54.1	644	1	YHJ9_YEAST
15	39	52.7	213	1	KAD_MYCCA
16	38	51.4	256	1	YD83_METJA
17	38	51.4	287	1	TRUB_AQUAE
18	38	51.4	572	1	HEMA_P13HT
19	37	50.0	191	1	Y096_HABIN
20	37	50.0	445	1	GNT1_HUMAN
21	37	50.0	447	1	GNT1_MOUSE
22	37	50.0	447	1	GNT1_RAT
23	37	50.0	490	1	Y032_BORBU
24	37	50.0	510	1	G6PD_ASPNG
25	37	50.0	511	1	G6PD_EMENI
26	37	50.0	548	1	YDD2_SCHPO
27	36	48.6	169	1	Y358_BUCAI
28	36	48.6	258	1	MIP_CHLPN
29	36	48.6	296	1	YD01_CLOAB
30	36	48.6	333	1	DPOB_XENLA
31	36	48.6	451	1	MURD_BACSU
32	36	48.6	461	1	NIFN_RHOCA
33	36	48.6	461	1	P19077 rhodobacter

RESULT 1

ID	TETX_CLOTE	STANDARD;	PRT;	1314 AA.
AC	P04958;			
DT	13-AUG-1987 (Rel. 05, Created)			
DT	13-AUG-1987 (Rel. 05, Last sequence update)			
DT	15-JUL-1999 (Rel. 38, Last annotation update)			
DE	Tetanus toxin precursor (EC 3.4.24.68) (Tentoxylisin).			
OS	Clostridium tetani.			
OG	Plasmid.			
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;			
OC	Clostridium			
OX	NCBI_TaxID=1513;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=87053814; PubMed=3536478;			
RA	Eisel U., Jarausch W., Goretzki K., Henschen A., Engels J.,			
RA	Weller U., Hudel M., Habermann E., Niemann H.;			
RT	"Tetanus toxin: primary structure, expression in E. coli, and			
RT	homology with botulinum toxins."			
RL	EMBO J. 5:2495-2502(1986).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CN3911;			
RX	MEDLINE=87040747; PubMed=3774547;			
RA	Fairweather N.F., Lyness V.A.;			
RT	"The complete nucleotide sequence of tetanus toxin."			
RL	Nucleic Acids Res. 14:7809-7812(1986).			
RN	[3]			
RP	SEQUENCE OF 742-1314 FROM N.A.			
RX	MEDLINE=86085672; PubMed=3510187;			
RA	Fairweather N.F., Lyness V.A., Pickard D.J., Allen G., Thomson R.O.;			
RT	"Cloning, nucleotide sequencing, and expression of tetanus toxin			
RT	fragment C in Escherichia coli."			
RL	J. Bacteriol. 165:21-27(1986).			
RN	[4]			
RP	PARTIAL SEQUENCE, AND DISULFIDE BONDS.			
RX	MEDLINE=90201034; PubMed=2108021;			
RA	Kriegelstein K., Henschen A., Weller U., Habermann E.;			
RT	"Arrangement of disulfide bridges and positions of sulphydryl groups			
RT	in tetanus toxin."			
RL	Eur. J. Biochem. 188:39-45(1990).			
RN	[5]			
RP	PARTIAL SEQUENCE.			
RX	MEDLINE=92037649; PubMed=1935979;			
RA	Kriegelstein K.G., Henschen A.H., Weller U., Habermann E.;			
RT	"Limited proteolysis of tetanus toxin. Relation to activity and			
RT	identification of cleavage sites."			
RL	Eur. J. Biochem. 202:41-51(1991).			
RN	[6]			
RP	IDENTIFICATION AS ZINC-PROTEASE.			
RX	MEDLINE=93010948; PubMed=1396558;			
RA	Schiavo G., Poulain B., Rossetto O., Benfenati F., Tauc L.,			
RA	Montecucco C.;			
RT	"Tetanus toxin is a zinc protein and its inhibition of			
RT	neurotransmitter release and protease activity depend on zinc."			

ALIGNMENTS

34	36	48.6	495	1	G6PD_PICJA
35	36	48.6	630	1	YND1_YEAST
36	36	48.6	774	1	RRP3_INCBE
37	36	48.6	774	1	RRP3_INCJJ
38	36	48.6	1630	1	MSPI_PLAFK
39	36	48.6	1639	1	MSPI_PLAFW
40	35	47.3	176	1	NU6C_SPIOL
41	35	47.3	200	1	ACD2_CIOAB
42	35	47.3	200	1	ACPD_SALTY
43	35	47.3	261	1	CABV_CHICK
44	35	47.3	294	1	CDD_ECOLI
45	35	47.3	321	1	YDG7_SCHPO

P11410	pichia jadi
P40009	saccharomyc
P21770	influenza c
P13877	influenza c
P04932	plasmodium
P04933	plasmodium
Q9m318	spinacia ol
Q97dd1	clostridium
Q8xf04	sallmonella
P04354	gallus gall
P13652	escherichia
Q10494	schizosacch

CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
 CC (Potential).
 CC -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
 CC kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
 CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M35727; AAA29715.1; -;
 CC EMBL; Y00087; CAA68280.1; -;
 CC EMBL; Z35326; CAA84555.1; -;
 CC PIR; S06286; S06286.
 CC InterPro: IPR000561; EGF-like.
 CC Pfam; PF00008; EGF; 1.
 CC Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
 CC Transmembrane; GPI-anchor.
 CC
 CC SIGNAL 1 19 POTENTIAL.
 CC CHAIN 20 1682 MEROZOITE SURFACE PROTEIN 1.
 CC TRANSMEM 1666 1682 MEMBRANE ANCHOR (POTENTIAL).
 CC CARBOHYD 233 233 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 462 462 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 528 528 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 599 599 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 785 785 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 881 881 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 901 901 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 947 947 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 1071 1071 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 1178 1178 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 1569 1569 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC SEQUENCE 1682 AA; 192462 MW; C82A1E159948CAD6 CRC64;
 CC
 CC Query Match 57.4%; Score 42.5; DB 1; Length 1682;
 CC Best Local Similarity 60.0%; Pred. No. 13;
 CC Matches 9; Conservative 4; Mismatches 1; Indels 1; Gaps 1;
 CC
 CC QY 1 QYIKANSKFI-GITE 14
 CC ::::|::|::|::|
 CC Db 983 QFVKSNSKVITGLTE 997
 CC
 CC RESULT 4
 CC MSP1_PLAFF STANDARD; PRT; 1701 AA.
 CC AC P13819;
 CC DT 01-JAN-1990 (Rel. 13, Created)
 CC DT 01-JAN-1990 (Rel. 13, Last sequence update)
 CC DT 01-OCT-1996 (Rel. 34, Last annotation update)
 CC DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
 CC DE (PMMSA).
 CC GN MSP-1.
 CC OS Plasmodium falciparum (isolate FC27 / Papua New Guinea).
 CC OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 CC OX NCBI_TaxID=5837;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RX MEDLINE=8814299; PubMed=2449612;
 CC RA Peterson M.G., Coppel R.L., McIntyre-P., Langford C.J., Woodrow G.,
 CC Brown G.V., Anders R.F., Kemp D.J.;
 CC "Variation in the precursor to the major merozoite surface antigens
 CC of Plasmodium falciparum";
 CC Mol. Biochem. Parasitol. 27:291-302(1988).
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
 CC (Potential).
 CC -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
 CC kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
 CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M19143; AAA29653.1; -;
 CC PIR; A54498; A54498.
 CC InterPro: IPR000561; EGF-like.
 CC Pfam; PF00008; EGF; 1.
 CC Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
 CC Transmembrane; GPI-anchor.
 CC
 CC SIGNAL 1 19 POTENTIAL.
 CC CHAIN 20 1701 MEROZOITE SURFACE PROTEIN 1.
 CC CARBOHYD 110 110 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 239 239 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 470 470 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 536 536 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 607 607 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 802 802 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 899 899 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 919 919 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 965 965 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 991 991 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 1089 1089 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 1196 1196 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 1588 1588 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC SEQUENCE 1701 AA; 193719 MW; 3920B75E73D38552 CRC64;
 CC
 CC Query Match 57.4%; Score 42.5; DB 1; Length 1701;
 CC Best Local Similarity 60.0%; Pred. No. 13;
 CC Matches 9; Conservative 4; Mismatches 1; Indels 1; Gaps 1;
 CC
 CC QY 1 QYIKANSKFI-GITE 14
 CC ::::|::|::|::|
 CC Db 1001 QFVKSNSKVITGLTE 1015
 CC
 CC RESULT 5
 CC MSP1_PLAFF STANDARD; PRT; 1701 AA.
 CC AC P08569;
 CC DT 01-AUG-1988 (Rel. 08, Created)
 CC DT 30-MAY-2000 (Rel. 39, Last sequence update)
 CC DT 30-MAY-2000 (Rel. 39, Last annotation update)
 CC DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
 CC DE (PMMSA) (P190).
 CC GN MSP-1.
 CC OS Plasmodium falciparum (isolate mad20 / Papua New Guinea).
 CC OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 CC OX NCBI_TaxID=70153;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RX MEDLINE=88011243; PubMed=3079521;
 CC RA Tanabe K., Mackay M., Goman M., Scaife J.G.;
 CC "Allelic dimorphism in a surface antigen gene of the malaria parasite
 CC Plasmodium falciparum";
 CC J. Mol. Biol. 195:273-287(1987).
 CC RN [2]
 CC RP REVISIONS TO 1403; 1569 AND 1629.
 CC RA Tanabe K.;
 CC RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 CC RN [3]
 CC RP SEQUENCE OF 1-115 FROM N.A.
 CC RX MEDLINE=86136024; PubMed=3004972;
 CC RA Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,
 CC Stunnenberg H., Bujard H.;
 CC "Polymorphism of the precursor for the major surface antigens of
 CC Plasmodium falciparum merozoites: studies at the genetic level";
 CC EMBO J. 4:3823-3829(1985).
 CC RL

CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
 CC (Potential).
 CC -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
 CC kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
 CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X05624; CAA29112.1; -
 CC PIR; A26868; A26868.
 CC PIR; B25120; B25120.
 CC InterPro; IPR000561; EGF-like.
 CC Pfam; PF00008; EGF; 1.
 CC Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
 CC Transmembrane; GPI-anchor.
 CC SIGNAL 1 19 POTENTIAL.
 CC CHAIN 20 1701 MEROZOITE SURFACE PROTEIN 1.
 CC CARBOHYD 110 110 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 239 239 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 470 470 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 536 536 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 607 607 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 802 802 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 899 899 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 919 919 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 965 965 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 991 991 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 1089 1089 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 1196 1196 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 1588 1588 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC SEQUENCE 1701 AA; 193768 MW; 3FC2EC59AF96EA98 CRC64;
 SQ
 Query Match 57.4%; Score 42.5; DB 1; Length 1701;
 Best Local Similarity 60.0%; Pred. No. 13;
 Matches 9; Conservative 4; Mismatches 1; Indels 1; Gaps 1;
 Qy 1 QYIKANSKFI-GITE 14
 Db 1001 QFVKSNSKVITGLTE 1015
 RESULT 6
 MSPI_PLAFC STANDARD; PRT; 1726 AA.
 AC P04934;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
 DE (PMMSA) (P195).
 GN MSP-1.
 OS Plasmodium falciparum (isolate Camp / Malaysia).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5835;
 RN [1]
 RP SEQUENCE OF 1-1103 FROM N.A.
 RX MEDLINE=86205236; PubMed=3517809;
 RA Weber J.L., Leininger W.M., Lyon J.A.;
 RT "Variation in the gene encoding a major merozoite surface antigen of
 RT the human malaria parasite Plasmodium falciparum.";
 RL Nucleic Acids Res. 14:3311-3323(1986).
 RN [2]
 RP SEQUENCE OF 1104-1726 FROM N.A.
 RX MEDLINE=88143999; PubMed=3278296;
 RA Weber J.L., Sim B.K.L., Lyon J.A., Wolff R.;
 RT "Merozoite surface protein sequence from the Camp strain of the human
 RT malaria parasite Plasmodium falciparum.";

RL Nucleic Acids Res. 16:1206-1206(1988).
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
 CC (Potential).
 CC -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
 CC kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
 CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X03831; CAA27446.1; -
 CC PIR; A23386; SAZQGM.
 CC InterPro; IPR000561; EGF-like.
 CC Pfam; PF00008; EGF; 1.
 CC Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
 CC Transmembrane; GPI-anchor.
 CC SIGNAL 1 19 POTENTIAL.
 CC CHAIN 20 1726 MEROZOITE SURFACE PROTEIN 1.
 CC CARBOHYD 133 133 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 272 272 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 501 501 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 567 567 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 638 638 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 827 827 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 924 924 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 944 944 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 990 990 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 1016 1016 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 1114 1114 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 1221 1221 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 1613 1613 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC SEQUENCE 1726 AA; 196197 MW; D08AD45FA352BCE3 CRC64;
 SQ
 Query Match 57.4%; Score 42.5; DB 1; Length 1726;
 Best Local Similarity 60.0%; Pred. No. 13;
 Matches 9; Conservative 4; Mismatches 1; Indels 1; Gaps 1;
 Qy 1 QYIKANSKFI-GITE 14
 Db 1026 QFVKSNSKVITGLTE 1040
 RESULT 7
 MSPI_PLAFC STANDARD; PRT; 1726 AA.
 AC P50435;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
 DE (PMMSA) (GP195).
 GN MSP-1.
 OS Plasmodium falciparum (isolate Palo Alto / Uganda).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=57270;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89005525; PubMed=3049134;
 RA Chang S.P., Kramer K.J., Yamaga K.M., Kato A., Case S.E.,
 RA Siddiqui W.A.;
 RT "Plasmodium falciparum: gene structure and hydrophathy profile of the
 RT major merozoite surface antigen (gp195) of the Uganda-Palo Alto
 RT isolate.";
 RL Exp. Parasitol. 67:1-11(1988).
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
 CC (Potential).
 CC -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
 CC kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF

```
CC -----
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M37213; AAA29611.1; -.
CC InterPro: IPR000561; EGF-like.
CC Pfam: PF00008; EGF; 1.
CC Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
CC Transmembrane; GPI-anchor.
CC -----
CC SIGNAL 1 19 POTENTIAL.
CC CHAIN 20 1726 MEROZOITE SURFACE PROTEIN 1.
CC CARBOHYD 133 133 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 272 272 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 501 501 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 567 567 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 638 638 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 827 827 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 924 924 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 944 944 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 990 990 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 1016 1016 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 1114 1114 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 1221 1221 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 1613 1613 N-LINKED (GLCNAC. .) (POTENTIAL).
CC SEQUENCE 1726 AA; 196174 MW; 5BS9CEFA2F9A026 CRC64;
CC -----
CC Query Match 57.4%; Score 42.5; DB 1; Length 1726;
CC Best Local Similarity 60.0%; Pred. No. 13;
CC Matches 9; Conservative 4; Mismatches 1; Indels 1; Gaps 1;
CC -----
CC QY 1 QYIKANSKFI-GITE 14
CC DB 1026 QFVKSNSKVITGLTE 1040
CC -----
CC RESULT 8
CC ID PYRC SERMA STANDARD; PRT; 204 AA.
CC AC Q9S3S1;
CC DT 16-OCT-2001 (Rel. 40, Created)
CC DT 16-OCT-2001 (Rel. 40, Last sequence update)
CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
CC DE Dihydroorotase (EC 3.5.2.3) (DHOase) (Fragment).
CC PYRC.
CC OS Serratia marcescens.
CC OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC OC Serratia.
CC OX NCBI_TaxID=615;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=SM6;
CC RA Berkmen M., Benedik M.J.;
CC RT "Dini inhibits transcription of Serratia marcescens nuclease.";
CC RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC CC -!- CATALYTIC ACTIVITY: (S)-dihydroorotate + H(2)O = N-carbamoyl-L-
CC aspartate.
CC CC -!- COFACTOR: THIS ENZYME TIGHTLY BINDS ONE ZINC ATOM PER CHAIN WHICH
CC IS REQUIRED FOR THE CATALYTIC MECHANISM. IT ALSO BINDS WEAKLY TO
CC TWO OTHERS. ZINCS WHICH ARE NOT ESSENTIAL FOR ACTIVITY (BY
CC SIMILARITY).
CC CC -!- PATHWAY: Pyrimidine biosynthesis; third step.
CC CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC CC -!- SIMILARITY: BELONGS TO THE DHOASE FAMILY. SUBFAMILY 1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

```
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF175466; AAD50307.1; ALT_INIT.
CC InterPro: IPR002195; Dihydroorotase.
CC DR PROSITE; PS00482; DIHYDROOROTASE 1; PARTIAL.
CC DR PROSITE; PS00483; DIHYDROOROTASE 2; 1.
CC KW Pyrimidine biosynthesis; Hydrolase; Zinc.
CC FT NON TER 1 1
CC SQ SEQUENCE 204 AA; 22899 MW; 9DB0B6C9B834B310 CRC64;
CC -----
CC Query Match 55.4%; Score 41; DB 1; Length 204;
CC Best Local Similarity 46.2%; Pred. No. 2; 9;
CC Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
CC -----
CC QY 1 QYIKANSKFI-GIT 13
CC DB 42 QYVQAGNRFLGAT 54
CC -----
CC RESULT 9
CC ID LE12 THETN STANDARD; PRT; 384 AA.
CC AC Q8RCF9;
CC DT 15-JUN-2002 (Rel. 41, Created)
CC DT 15-JUN-2002 (Rel. 41, Last sequence update)
CC DT 15-JUN-2002 (Rel. 41, Last annotation update)
CC DE 2-isopropylmalate synthase 2 (EC 4.1.3.12) (Alpha-isopropylmalate
CC synthase 2) (Alpha-IPM synthetase 2).
CC GN LEU2A OR TTE0472.
CC OS Thermoanaerobacter tengcongensis.
CC OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
CC OC Thermoanaerobacteriaceae; Thermoanaerobacter.
CC OX NCBI_TaxID=119072;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=MB4 / JCM 11007;
CC RX MEDLINE=21992816; PubMed=11997336;
CC RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
CC Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
CC Tan H., Chen R., Wang J., Yu J., Yang H.;
CC RT "A complete sequence of T. tengcongensis genome.";
CC RL Genome Res. 12:689-700(2002).
CC CC -!- FUNCTION: Catalyzes the condensation of the acetyl group of
CC acetyl-CoA with 3-methyl-2-oxobutanoate (2-oxoisovalerate) to form
CC 3-carboxy-3-hydroxy-4-methylpentanoate (2-isopropylmalate).
CC CC -!- CATALYTIC ACTIVITY: 3-carboxy-3-hydroxy-4-methylpentanoate + CoA =
CC acetyl-CoA + 3-methyl-2-oxobutanoate + H(2)O.
CC CC -!- PATHWAY: Leucine biosynthesis; first step.
CC CC -!- SUBUNIT: Homotetramer (By similarity).
CC CC -!- SIMILARITY: BELONGS TO THE ALPHA-IPM SYNTHETASE / HOMOCITRATE
CC SYNTHASE FAMILY. LEU2A 1 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AB013018; AA023753.1; -.
CC DR PROSITE; PS00815; AIPM HOMOCIT SYNTH 1; 1.
CC DR PROSITE; PS00816; AIPM HOMOCIT SYNTH 2; 1.
CC KW Leucine biosynthesis; Lyase; Complete proteome.
CC SQ SEQUENCE 384 AA; 42404 MW; 095310F2C0B4A4DD CRC64;
CC -----
CC Query Match 55.4%; Score 41; DB 1; Length 384;
CC Best Local Similarity 54.5%; Pred. No. 5; 5;
CC Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
```

```

OY 3 IKANSKFIGIT 13
DB 214 VKAGAKFVGVT 224

RESULT 10
ID V120 HSVSA STANDARD; PRT; 899 AA.
AC Q01055;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Capsid assembly protein 63.
GN 63 OR EERRF1.
OS Herpesvirus saimiri (strain 11).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=10383;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92333688; PubMed=1321287;
RA Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Biesinger B.,
RA Newman C., Wittmann S., Craxton M.A., Coleman H., Fleckenstein B.,
RA Honess R.W.;
RA "Primary structure of the herpesvirus saimiri genome.";
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92230226; PubMed=1314457;
RA Nicholas J., Cameron K.R., Coleman H., Newman C., Honess R.W.;
RA "Analysis of nucleotide sequence of the rightmost 43 kbp of
RA herpesvirus saimiri (HVS) L-DNA: General conservation of genetic
RA organization between HVS and Epstein-Barr virus.";
RL Virology 188:296-310(1992).
CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL37,
CC HSV-1 23, EBV BOLFI, VZV 21, HVS-1 63, AND HCMV UL47.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X64346; CAA45686.1; -.
DR EMBL; M86409; AAA46139.1; -.
DR PIR; G36812; G36812.
KW Capsid assembly.
SQ SEQUENCE 899 AA; 103350 MW; F1429B3770A2885E CRC64;

Query Match 55.4%; Score 41; DB 1; Length 899;
Best Local Similarity 50.0%; Pred. No. 13;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 1 QYIKANSKFIGITE 14
DB 124 QYITSNATFGTGLSE 137

RESULT 11
ACPD HAEN
ID ACPD HAEN STANDARD; PRT; 194 AA.
AC P43013;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable acyl carrier protein phosphodiesterase (EC 3.1.4.14) (ACP
DE phosphodiesterase).
GN ACPD:OR H11366
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.

OY 3 IKANSKFIGIT 13
DB 214 VKAGAKFVGVT 224

RESULT 12
PEFI LACLC
ID PEFI LACLC STANDARD; PRT; 601 AA.
AC P54124; P94880;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Oligoendopeptidase F, plasmid (EC 3.4.24.-).
GN PEPF1 OR PEPF.
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OG Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1359;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=NCDO 763;
RX MEDLINE=95096044; PubMed=7798200;
RA Monnet V., Nardi M., Chopin A., Chopin M.-C., Gripon J.-C.;
RA "Biochemical and genetic characterization of PepF, an oligopeptidase

```


Search completed: July 22, 2003, 08:06:39
Job time : 2.39504 secs

```
DR SGD; S0001081; YHR039C.
DR InterPro; IPR002086; Aldenhyde_dehydr.
DR Pfam; PF00171; aldedh; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Hypothetical protein; Oxidoreductase.
FT ACT SITE 354 354 BY SIMILARITY.
FT ACT SITE 389 389 BY SIMILARITY.
SQ SEQUENCE 644 AA; 71320 MW; 54DADDAB2A16D4D CRC64;

Query Match 54.1%; Score 40; DB 1; Length 644;
Best Local Similarity 60.0%; Pred. No. 14;
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 QYIKANSKFIGITEL 15
Db 38 QIQNQKLIGITTL 52

RESULT 15
KAD_MYCCA STANDARD; PRT; 213 AA...
AC P10251;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Adenylate kinase (EC 2.7.4.3) (ATP-AMP transphosphorylase).
GN ADK.
OS Mycoplasma capricolum.
OC Bacteria; Firmicutes; Mollicutes; Entomoplasmatales;
OC Entomoplasmataceae.
OX NCBI_TaxID=2095;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 27343 / KID;
RA MEDLINE=88142549; PubMed=3481422;
RX Okubo S., Muto A., Kawachi Y., Yamao F., Osawa S.;
RT "The ribosomal protein gene cluster of Mycoplasma capricolum.";
RL Mol. Gen. Genet. 210:314-322(1987).
CC -1- FUNCTION: THIS SMALL UBIQUITOUS ENZYME IS ESSENTIAL FOR
CC MAINTENANCE AND CELL GROWTH.
CC -1- CATALYTIC ACTIVITY: ATP + AMP = ADP + ADP.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE ADENYLATE KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X06414; CAA29724.1; -.
DR PIR; S02851; KIYMC.
DR HSSP; P27142; 1ZIN.
DR InterPro; IPR000850; Adenylate_kin.
DR Pfam; PF00406; adenylatekinase; 1.
DR PRINTS; PR00094; ADENYLTKINASE.
DR PRODOM; PD000657; Adenylate_kin; 1.
DR PROSITE; PS00113; ADENYLATE_KINASE; 1.
KW Transferase; Kinase; ATP-binding.
FT NP BIND 7 15 ATP (BY SIMILARITY).
SQ SEQUENCE 213 AA; 24616 MW; FCFPA426B6F92E16 CRC64;

Query Match 52.7%; Score 39; DB 1; Length 213;
Best Local Similarity 72.7%; Pred. No. 7;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 YIKANSKFIGI 12
Db 185 YFKTNSKPIEI 193
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:05:43 ; Search time 0.55306 Seconds
(without alignments)
1574.881 Million cell updates/sec

Title: US-09-806-703A-14.

Perfect score: 112

Sequence: 1 FNNFTVFWLRVPKVSASHLE 21

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	112	100.0	1314	1	TETX_CLOTE
2	61	54.5	1274	1	BXF_CLOBO
3	61	54.5	1296	1	BXG_CLOBO
4	59	52.7	1295	1	BXA1_CLOBO
5	58	51.8	1290	1	BXB_CLOBO
6	57	50.9	1051	1	VP2_AHSV6
7	56	50.0	1250	1	BXE_CLOBO
8	56	50.0	1250	1	BXE_CLOBO
9	56	50.0	1295	1	BXA2_CLOBO
10	52	46.4	449	1	VNSS_INSVN
11	52	46.4	464	1	VNSS_TSNV1
12	52	46.4	467	1	VNSS_TSNV1
13	51	45.5	1196	1	BXCX_CLOBO
14	48	42.9	1290	1	BXC1_CLOBO
15	47.5	42.4	1276	1	BXD_CLOBO
16	47	42.0	760	1	AMY_CLOAB
17	46	41.1	1162	1	BXEN_CLOBO
18	46	41.1	1162	1	BXEN_CLOBO
19	45	40.2	639	1	CGMA_RHIME
20	44	39.3	241	1	CRTA_RHOCA
21	43	38.4	152	1	YOBID_ECOLI
22	43	38.4	867	1	RRPO_BYDV1
23	42	37.5	208	1	FTSQ_STRGR
24	42	37.5	789	1	A12M_YEAST
25	42	37.5	1337	1	YDMS_SCHPO
26	42	37.5	1539	1	SMCY_HUMAN
27	42	37.5	1548	1	SMCY_MOUSE
28	42	37.5	1554	1	SMCX_MOUSE
29	42	37.5	1560	1	SMCX_HUMAN
30	41.5	37.1	237	1	YBSO_YEAST
31	41.5	37.1	728	1	YJ88_YEAST
32	41	36.6	352	1	CITC_ECOLI
33	40.5	36.2	230	1	FIMC_SALTY

RESULT 1

ID	TETX_CLOTE	STANDARD;	PRT;	1314 AA.
AC	P04958;			
DT	13-AUG-1987 (Rel. 05, Created)			
DT	13-AUG-1987 (Rel. 05, Last sequence update)			
DT	15-JUL-1999 (Rel. 38, Last annotation update)			
DE	Tetanus toxin precursor (EC 3.4.24.68) (Tentoxylisin).			
OS	Clostridium tetani.			
OG	Plasmid.			
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;			
OC	Clostridium.			
OX	NCBI_TaxID=1513;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=87053814; PubMed=3536478;			
RA	Eisel U., Jarausch W., Goretzki K., Henschen A., Engels J.,			
RA	Weller U., Hudel M., Habermann E., Niemann H.;			
RT	"Tetanus toxin: primary structure, expression in E. coli, and			
RT	homology with botulinum toxins.";			
RL	EMBO J. 5:2495-2502(1986).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CN3911;			
RX	MEDLINE=87040747; PubMed=3774547;			
RA	Fairweather N.P., Lyness V.A.;			
RT	"The complete nucleotide sequence of tetanus toxin.";			
RL	Nucleic Acids Res. 14:7809-7812(1986).			
RN	[3]			
RP	SEQUENCE OF 742-1314 FROM N.A.			
RX	MEDLINE=86085672; PubMed=3510187;			
RA	Fairweather N.P., Lyness V.A., Pickard D.J., Allen G., Thomson R.O.;			
RT	"Cloning, nucleotide sequencing, and expression of tetanus toxin			
RT	fragment C in Escherichia coli.";			
RL	J. Bacteriol. 165:21-27(1986).			
RN	[4]			
RP	PARTIAL SEQUENCE, AND DISULFIDE BONDS.			
RX	MEDLINE=92037649; PubMed=2108021;			
RA	Kriegelstein K., Henschen A., Weller U., Habermann E.;			
RT	"Arrangement of disulfide bridges and positions of sulphhydryl groups			
RT	in tetanus toxin.";			
RL	Eur. J. Biochem. 188:39-45(1990).			
RN	[5]			
RP	PARTIAL SEQUENCE.			
RX	MEDLINE=93010948; PubMed=1396558;			
RA	Kriegelstein K.G., Henschen A.H., Weller U., Habermann E.;			
RT	"Limited proteolysis of tetanus toxin. Relation to activity and			
RT	identification of cleavage sites.";			
RL	Eur. J. Biochem. 202:41-51(1991).			
RN	[6]			
RP	IDENTIFICATION AS ZINC-PROTEASE.			
RX	MEDLINE=93010948; PubMed=1396558;			
RA	Schiavo G., Poulain B., Rossetto O., Benfenati F., Tauc L.,			
RA	Montecucco C.;			
RT	"Tetanus toxin is a zinc protein and its inhibition of			
RT	neurotransmitter release and protease activity depend on zinc.";			

Q09786 schizosacch
P92817 paralichthy
Q04276 saccharomyc
P47706 mycoplasma
O74456 schizosacch
P58952 drosophila
P37604 salmonella
P47049 saccharomyc
Q8tw29 methanopyru
P03470 influenza a
P03468 influenza a
P06819 influenza a

34 40.5 36.2 535 1 YA98 SCHPO
35 40 35.7 116 1 NU3M PAROL
36 40 35.7 118 1 YMX2 YEAST
37 40 35.7 269 1 PARA MYCGE
38 40 35.7 288 1 PER1 SCHPO
39 40 35.7 383 1 G22C DROME
40 40 35.7 390 1 DACD SALTY
41 40 35.7 396 1 YJEB YEAST
42 40 35.7 402 1 LE22 METKA
43 40 35.7 453 1 NRAM IAWIL
44 40 35.7 454 1 NRAM IAPUE
45 40 35.7 469 1 NRAM IAPAR

ALIGNMENTS

EMBO J. 11:3577-3583(1992).
 [7]
 IDENTIFICATION OF SUBSTRATE.
 MEDLINE=93063293; PubMed=1331807;
 Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,
 Dasgupta B.R., Montecucco C.;
 "Tetanus and botulinum-B neurotoxins block neurotransmitter release
 by proteolytic cleavage of synaptobrevin.";
 Nature 359:832-835(1992).
 [8]
 X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 874-1314.
 MEDLINE=97475217; PubMed=9334741;
 Unland T.C., Wingert L.M., Swaminathan S., Furey W.F., Schmidt J.J.,
 Sax M.;
 "Structure of the receptor binding fragment HC of tetanus
 neurotoxin.";
 Nat. Struct. Biol. 4:788-792(1997).
 CC -1- FUNCTION: TETANUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
 RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
 AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
 WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
 INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
 ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 76-GLN-|-PHE-77
 BOND OF SYNAPTOSOMAL-2.
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 76-GLN-|-PHE-77 BOND IN
 SYNAPTOSOMAL-2.
 CC -1- SUBUNIT: THE PRECURSOR POLYPEPTIDE IS SUBSEQUENTLY CLEAVED TO
 YIELD SUBCHAINS L AND H. THESE REMAIN LINKED BY A DISULFIDE BRIDGE
 AND ARE NON-TOXIC AFTER SEPARATION.
 CC -1- MISCELLANEOUS: THE C-TERMINAL OF THE HEAVY CHAIN BINDS TO
 GANGLIOSIDE RECEPTORS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X04436; CAA28033.1; -
 DR EMBL; M12739; AAA23282.1; -
 DR EMBL; X06214; CAA29564.1; -
 DR PIR; A25689; BTCLTN
 DR PDB; 1AF9; 29-APR-98.
 DR PDB; 1ABD; 14-OCT-98.
 DR MEROPS; M27.001; -
 DR InterPro; IPR000395; Bontoxilysin.
 DR Pfam; PF01742; Peptidase_M27; 1.
 DR PRINTS; PR00760; BONTOKILYSIN.
 DR PRODOM; PD001963; Bontoxilysin; 1.
 DR PROSITE; PS00142; ZINC PROTEASE; 1.
 KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc; Plasmid;
 KM 3D-structure.
 FT INIT MET 0 0
 FT CHAIN 1 456 TETANUS TOXIN LIGHT CHAIN.
 FT CHAIN 457 1314 TETANUS TOXIN HEAVY CHAIN.
 FT METAL 232 232 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT SITE 233 233 BY SIMILARITY.
 FT METAL 236 236 ZINC (CATALYTIC) (BY SIMILARITY).
 FT TRANSMEM 226 246 POTENTIAL.
 FT TRANSMEM 669 689 INTERCHAIN.
 FT DISULFID 438 466
 FT DISULFID 1076 1092
 SQ SEQUENCE 1314 AA; 150550 MW; 134C3657133EP81D CRC64;
 Query Match 100.0%; Score 112; DB 1; Length 1314;
 Best Local Similarity 100.0%; Pred. No. 1.8e-10;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 FNNFTVSFVLRVVKVSASHLE 21

Db 946 FNNFTVSFVLRVVKVSASHLE 966
 |||||
 RESULT 2
 BXF_CLOBO
 ID BXF_CLOBO STANDARD; PRT; 1274 AA.
 AC P30996;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Botulinum neurotoxin type F precursor (EC 3.4.24.69) (BoNT/F)
 DE (Bontoxilysin F).
 GN BOTF.
 OS Clostridium botulinum.
 OS Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1491;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 23387;
 RX MEDLINE=93012902; PubMed=1398040;
 RA East A.K., Richardson P.T., Allaway D., Collins M.D.,
 RA Roberts T.A., Thompson D.E.;
 RT "Sequence of the gene encoding type F neurotoxin of Clostridium
 botulinum";
 RT FEMS Microbiol. Lett. 75:225-230(1992).
 RN [2]
 RP SEQUENCE OF 1-64 FROM N.A.
 RC STRAIN=Hobbs FT10;
 RX MEDLINE=94297488; PubMed=7764998;
 RA East A.K., Collins M.D.;
 RT "Conserved structure of genes encoding components of botulinum
 neurotoxin complex M and the sequence of the gene coding for the
 nontoxic component in nonproteolytic Clostridium botulinum type F";
 Curr. Microbiol. 29:69-77(1994).
 RN [3]
 RP SEQUENCE OF 634-1002 FROM N.A.
 RX MEDLINE=94013372; PubMed=8408542;
 RA Campbell K., East A.K., Collins M.D.;
 RT "Gene probes for identification of the botulinum neurotoxin gene and
 specific identification of neurotoxin types B, E, and F";
 J. Clin. Microbiol. 31:2255-2262(1993).
 RN [4]
 RP IDENTIFICATION OF SUBSTRATE.
 RX MEDLINE=94230352; PubMed=8175689;
 RA Yamaeaki S., Baumeister A., Binz T., Blas J., Link E., Cornille F.,
 RA Roques B., Fykse E.M., Suedhof T.C., Jahn R., Niemann H.;
 RT "Cleavage of members of the synaptobrevin/VAMP family by types D and
 F botulinum neurotoxins and tetanus toxin";
 J. Biol. Chem. 269:12764-12772(1994).
 CC -1- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
 RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
 AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
 WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
 INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
 ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 58-GLN-|-LYS-59
 BOND OF SYNAPTOSOMAL-1 AND -2.
 CC -1- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
 neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
 detected action on small molecule substrates.
 CC -1- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
 HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
 WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL
 FORMATION AND TOXIN BINDING, RESPECTIVELY.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
 BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL; M92906; AAA23263.1; -;
 DR DR EMBL; S71676; AAC60475.1; -;
 DR DR EMBL; X70820; CAA50151.1; -;
 DR DR EMBL; X70816; CAA50147.1; -;
 DR DR HSSP; P10845; 3BTA.
 DR DR MEROPS; M27.002; -;
 DR DR InterPro; IPR000395; Bontoxilysin.
 DR DR InterPro; IPR000130; Zn MTpeptdase.
 DR DR Pfam; PF01742; Peptidase M27; 1.
 DR DR PRINTS; PR00760; BONTOXILYSIN.
 DR DR PRODOM; PD001963; Bontoxilysin; 1.
 DR DR PROSITE; PS00142; ZINC PROTEASE; 1.
 DR DR Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
 FT CHAIN 1 436 BOTULINUM NEUROTOXIN F, LIGHT-CHAIN.
 FT CHAIN 437 1274 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 227 227 BY SIMILARITY.
 FT ACT SITE 228 228 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 231 231 ZINC (CATALYTIC) (BY SIMILARITY).
 FT DISULFID 429 445 INTERCHAIN (PROBABLE).
 SQ SEQUENCE 1274 AA; 146709 MW; 5B99756A7438B921 CRC64;

Query Match 54.5%; Score 61; DB 1; Length 1274;
 Best Local Similarity 57.1%; Pred. No. 0.049;
 Matches 8; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 FNNFTVSFWLRVPK 14

DB 930 YQNFSSFWVRIPK 943

RESULT 3

BXG_CLOBO STANDARD; PRT; 1296 AA.
 AC Q60393;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Botulinum neurotoxin type G precursor (EC 3.4.24.69) (BONT/G)
 DE (Bontoxilysin G).
 DE BOTG.
 GN Clostridium botulinum.
 OS Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1491;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=113 / 30;
 RX MEDLINE=94092745; PubMed=8268233;
 RA Campbell K., Collins M.D., East A.K.;
 RT "Nucleotide sequence of the gene coding for Clostridium botulinum
 RT (Clostridium argentinense) type G neurotoxin: genealogical comparison
 RT with other clostridial neurotoxins."
 RL Biochim. Biophys. Acta 1216:487-491(1993).
 CC -1- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
 CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
 CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
 CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
 CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
 CC ENDOPEPTIDASE.
 CC -1- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
 CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
 CC detected action on small molecule substrates.
 CC -1- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
 CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
 CC WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL
 CC FORMATION AND TOXIN BINDING, RESPECTIVELY.
 CC -1- SUBCELLULAR LOCATION: Secreted (by similarity).
 CC -1- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF

CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; X74162; CAA52275.1; -;
 DR DR HSSP; P10845; 3BTA.
 DR DR MEROPS; M27.002; -;
 DR DR InterPro; IPR000395; Bontoxilysin.
 DR DR InterPro; IPR000130; Zn MTpeptdase.
 DR DR Pfam; PF01742; Peptidase M27; 1.
 DR DR PRINTS; PR00760; BONTOXILYSIN.
 DR DR PRODOM; PD001963; Bontoxilysin; 1.
 DR DR PROSITE; PS00142; ZINC PROTEASE; 1.
 DR DR Neurotoxin; Hydrolase; Metalloprotease; Zinc.
 FT INIT MET 0 0 BY SIMILARITY.
 FT CHAIN 1 441 BOTULINUM NEUROTOXIN G, LIGHT-CHAIN.
 FT CHAIN 442 1296 BOTULINUM NEUROTOXIN G, HEAVY-CHAIN.
 FT METAL 229 229 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT SITE 230 230 BY SIMILARITY.
 FT METAL 233 233 ZINC (CATALYTIC) (BY SIMILARITY).
 FT DISULFID 435 449 INTERCHAIN (PROBABLE).
 SQ SEQUENCE 1296 AA; 149013 MW; DC8E47E15F665C31 CRC64;

Query Match 54.5%; Score 61; DB 1; Length 1296;
 Best Local Similarity 38.1%; Pred. No. 0.049;
 Matches 8; Conservative 9; Mismatches 4; Indels 0; Gaps 0;

QY 1 FNNFTVSFWLRVPKVSASHLE 21

DB 929 FDNFSINFWRTPKYNNDIQ 949

RESULT 4

BXAL_CLOBO STANDARD; PRT; 1295 AA.
 ID EXAL_CLOBO
 AC P10845; P18639; P01561;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BONT/A)
 DE (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-
 DE chain; Botulinum neurotoxin A, heavy-chain].
 GN BOTA OR BNA OR ATX.
 OS Clostridium botulinum.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1491;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCTC 2916;
 RX MEDLINE=90235864; PubMed=2185020;
 RA Thompson D.E., Brehm J.K., Oultram J.D., Swinfield T.-J.,
 RA Shone C.C., Atkinson T., Melling J., Minton N.P.;
 RT "The complete amino acid sequence of the Clostridium botulinum type A
 RT neurotoxin, deduced by nucleotide sequence analysis of the encoding
 RT gene."
 RL Eur. J. Biochem. 189:73-81(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=62A;
 RX MEDLINE=90264400; PubMed=2160960;
 RA Binz B., Kuaronzo H., Wille M., Frevent J., Wernars K., Niemann H.;
 RT "The complete sequence of botulinum neurotoxin type A and comparison
 RT with other clostridial neurotoxins."
 RL J. Biol. Chem. 265:9153-9158(1990).
 RN [3]

RP SEQUENCE OF 1-65 FROM N.A.
RC STRAIN=62A;
RX MEDLINE=97016817; PubMed=8863443;
RA East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
RT "Organization and phylogenetic interrelationships of genes encoding
RT components of the botulinum toxin complex in proteolytic Clostridium
RT botulinum types A, B, and F: evidence of chimeric sequences in the
RT gene encoding the nontoxic nonhemagglutinin component.";
RL Int. J. Syst. Bacteriol. 46:1105-1112(1996).
RN [4]
RP SEQUENCE OF 1-34 FROM N.A.
RC STRAIN=Hall;
RX MEDLINE=89350959; PubMed=2669749;
RA Betley M.J., Somers E., Dasgupta B.R.;
RT "Characterization of botulinum type A neurotoxin gene: delineation of
RT the N-terminal encoding region.";
RL Biochem. Biophys. Res. Commun. 162:1388-1395(1989).
RN [5]
RP SEQUENCE OF 1-18 FROM N.A.
RC STRAIN=Type A NIH;
RX MEDLINE=96096783; PubMed=8521962;
RA Fujita R., Fujinaga Y., Inoue K., Nakajima H., Kumon H., Oguma K.;
RT "Molecular characterization of two forms of nontoxic-nonhemagglutinin
RT components of Clostridium botulinum type A progenitor toxins.";
RL FEBS Lett. 376:41-44(1995).
RN [6]
RP SEQUENCE OF 1-16.
RX MEDLINE=84178501; PubMed=6370252;
RA Schmidt J.J., Sartymoorthy V., Dasgupta B.R.;
RT "Partial amino acid sequence of the heavy and light chains of
RT botulinum neurotoxin type A.";
RL Biochem. Biophys. Res. Commun. 119:900-904(1984).
RN [7]
RP SEQUENCE OF 1-46.
RA Dasgupta B.R., Foley J., Niece R.;
RT "Partial sequence of the light chain of botulinum neurotoxin type A.";
RL Biochemistry 26:4162-4162(1987).
RN [8]
RP SEQUENCE OF 1-5 AND 444-456.
RX MEDLINE=91120847; PubMed=2126206;
RA Dasgupta B.R., Dekleva M.B.;
RT "Botulinum neurotoxin type A: sequence of amino acids at the
RT N-terminus and around the nicking site.";
RL Biochimie 72:661-664(1990).
RN [9]
RP SEQUENCE OF 448-464 AND 872-895.
RX MEDLINE=89044662; PubMed=3178218;
RA Sartymoorthy V., Dasgupta B.R., Foley J., Niece R.L.;
RT "Botulinum neurotoxin type A: cleavage of the heavy chain into two
RT halves and their partial sequences.";
RL Arch. Biochem. Biophys. 266:142-151(1988).
RN [10]
RP SEQUENCE OF 448-482.
RX MEDLINE=85285016; PubMed=3896784;
RA Shone C.C., Hambleton P., Melling J.;
RT "Inactivation of Clostridium botulinum type A neurotoxin by trypsin
RT and purification of two tryptic fragments. Proteolytic action near
RT the COOH-terminus of the heavy subunit destroys toxin-binding
RT activity.";
RL Eur. J. Biochem. 151:75-82(1985).
RN [11]
RP IDENTIFICATION OF SUBSTRATE.
RX MEDLINE=94063091; PubMed=8243676;
RA Schiavo G., Santucci A., Dasgupta B.R., Mehta P.P., Jontes J.,
RT Benfenati F., Wilson M.C., Montecucco C.;
RT "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
RT COOH-terminal peptide bonds.";
RL FEBS Lett. 335:99-103(1993).
RN [12]
RP IDENTIFICATION OF SUBSTRATE.
RX MEDLINE=94124495; PubMed=8294407;
RA Binz T., Blasi J., Yasasaki S., Baumeister A., Link E., Suedhof T.C.,
RA Jahn R., Niemann H.;

RT "Proteolysis of SNAP-25 by types E and A botulin neurotoxins.";
RL J. Biol. Chem. 269:1617-1620(1994).
RN [13]
RP MUTAGENESIS OF GLU-261; PHE-265 AND TYR-365.
RX MEDLINE=21556941; PubMed=11700044;
RA Rigoni M., Caccin P., Johnson E.A., Montecucco C., Rossetto O.;
RT "Site-directed mutagenesis identifies active-site residues of the
RT light chain of botulinum neurotoxin type A.";
RL Biochem. Biophys. Res. Commun. 288:1231-1237(2001).
RN [14]
RP X-RAY CRYSTALLOGRAPHY (3.3 ANGSTROMS).
RX MEDLINE=98455071; PubMed=9783750;
RA Lacy D.B., Tepp W., Cohen A.C., Dasgupta B.R., Stevens R.C.;
RT "Crystal structure of botulinum neurotoxin type A and implications
RT for toxicity.";
RL Nat. Struct. Biol. 5:898-902(1998).
CC -I- FUNCTION: Inhibits acetylcholine release. The botulinum toxin
CC binds with high affinity to peripheral neuronal presynaptic
CC membrane, is then internalized by receptor-mediated endocytosis.
CC The C-terminus of the heavy chain (H) is responsible for the
CC adherence of the toxin to the cell surface while the N-terminus
CC mediates transport of the light chain from the endocytic vesicle
CC to the cytosol. After translocation, the light chain (L)
CC hydrolyzes the 197-Gln-Arg-198 bond in SNAP-25, thereby blocking
CC neurotransmitter release. Inhibition of acetylcholine release
CC results in flaccid paralysis, with frequent heart or respiratory
CC failure.
CC -I- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC detected action on small molecule substrates.
CC -I- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC heavy chain (H).
CC -I- SUBCELLULAR LOCATION: Secreted.
CC -I- PHARMACEUTICAL: Available under the name BOTOX (Allergan) for
CC the treatment of strabismus and blepharospasm associated with
CC dystonia and cervical dystonia. Also used for the treatment of
CC hemifacial spasm and a number of other neurological disorders
CC characterized by abnormal muscle contraction.
CC -I- MISCELLANEOUS: There are seven antigenically distinct forms of
CC botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC -I- DATABASE: NAME=BOTOX product information web site;
CC WWW="http://www.botox.com/index.jsp?hp&productinfo".
CC -I- DATABASE: NAME=Protein Spotlight;
CC NOTE=Issue 19 of February 2002;
CC WWW="http://www.expasy.org/spotlight/articles/spotlt019.html".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X52066; CA336289.1; -
DR EMBL; M30196; AAA23262.1; -
DR EMBL; X92973; CAA83551.1; -
DR EMBL; D67030; BAA11051.1; -
DR EMBL; M27892; AAA23269.1; -
DR PIR; A35294; BTCLAB.
DR PIR; S09492; S09492.
DR PDB; 3BTA; 01-OCT-99.
DR MEROPS; M27.002; -
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR000130; Zn_MTPeptidse.
DR Pfam; PF01742; Peptidase M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc;
KW Pharmaceutical; 3D-structure.
FT INIT_MET 0

```

FT CHAIN 1 447 BOTULINUM NEUROTOXIN A, LIGHT-CHAIN.
FT CHAIN 448 1295 BOTULINUM NEUROTOXIN A, HEAVY-CHAIN.
FT METAL 222 ZINC (CATALYTIC).
FT ACT_SITE 223
FT METAL 226 ZINC (CATALYTIC).
FT METAL 261 ZINC (CATALYTIC).
FT DISULFID 429 453 INTERCHAIN.
FT TRANSMEM 626 646 POTENTIAL.
FT TRANSMEM 655 675 POTENTIAL.
FT VARIANT 26 26 V -> A.
FT MUTAGEN 261 261 E->A: DRASTIC DECREASE IN ENZYMIC ACTIVITY.
FT MUTAGEN 265 265 F->A: DECREASE IN ENZYMIC ACTIVITY.
FT MUTAGEN 365 365 Y->A: DECREASE IN ENZYMIC ACTIVITY.
FT CONFLICT 1 1 P -> Q (IN REF. 1).
FT CONFLICT 479 479 E -> P (IN REF. 9).
FT CONFLICT 875 875 T -> L (IN REF. 8).
FT CONFLICT 891 891 S -> K (IN REF. 8).
SQ SEQUENCE 1295 AA; 149322 MW; 858342F754862579 CRC64;

Query Match 52.7%; Score 59; DB 1; Length 1295;
Best Local Similarity 57.1%; Pred. No. 0.11;
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 FNNFTVSFWLRVPK 14
Db 937 YENFSTFWLRVPK 950

RESULT 5
BXB CLOBO
ID AC BXB CLOBO STANDARD; PRT; 1290 AA.
AD P10844; P10843;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Botulinum neurotoxin type B precursor (EC 3.4.24.69) (BoNT/B)
DE (Bontoxilysin B).
GN BOTB.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
[1]
RN RP SEQUENCE FROM N.A.
RX MEDLINE=92384550; PubMed=1514783;
RA Whelan S.M., Elmore M.J., Bodsworth N.J., Brehm J.K., Atkinson T.,
RA Minton N.P.;
RT "Molecular cloning of the Clostridium botulinum structural gene
RT encoding the type B neurotoxin and determination of its entire
RT nucleotide sequence."
RL Appl. Environ. Microbiol. 58:2345-2354 (1992).
RN [2]
RP SEQUENCE OF 35-245 FROM N.A.
RC STRAIN=NTC 7273;
RA Szabo E.A., Pemberton J.M., Desmarchelier P.M.;
RL Submitted (APR-1992) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 633-993 FROM N.A.
RC STRAIN=NTC 7273;
RX MEDLINE=94013372; PubMed=8408542;
RA Campbell K., East A.K., Collins M.D.;
RT "Gene probes for identification of the botulinum neurotoxin gene and
RT specific identification of neurotoxin types B, E, and F."
RL J. Clin. Microbiol. 31:2255-2262 (1993).
RN [4]
RP SEQUENCE OF 1-44 AND 441-466.
RC STRAIN=657;
RX MEDLINE=89000987; PubMed=3139097;
RA Dasgupta B.R., Datta A.;
RT "Botulinum neurotoxin type B (strain 657): partial sequence and
RT similarity with tetanus toxin."

```

```

RL Biochimie 70:811-817 (1988).
RN [5]
RP SEQUENCE OF 1-16 AND 441-458.
RC STRAIN=OKRA;
RX MEDLINE=85197963; PubMed=3888113;
RA Schmidt J.J., Sathyanarayanan V., Dasgupta B.R.;
RT "Partial amino acid sequences of botulinum neurotoxins types B and
RT E."
RN Arch. Biochem. Biophys. 238:544-548 (1985).
RL [6]
RP IDENTIFICATION AS ZINC-PROTEASE.
RX MEDLINE=93054694; PubMed=1429690;
RA Schiavo G., Rossetto O., Santucci A., Dasgupta B.R., Montecucco C.;
RT "Botulinum neurotoxins are zinc proteins."
RL J. Biol. Chem. 267:23479-23483 (1992).
RN [7]
RP IDENTIFICATION OF SUBSTRATE.
RX MEDLINE=93063293; PubMed=1331807;
RA Schiavo B.R., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,
RA Dasgupta B.R., Montecucco C.;
RT "Tetanus and botulinum-B neurotoxins block neurotransmitter release
RT by proteolytic cleavage of synaptobrevin."
RL Nature 359:832-835 (1992).
CC -1- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE THAT CLEAVES THE 76-GLN-|-PHE-77 BOND OF
CC SYNAPTOSOMAL-2.
CC -1- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC detected action on small molecule substrates.
CC -1- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
CC WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL
CC FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M81186; AAA23211.1; -
CC EMBL; Z11934; CAA77591.1; -
CC EMBL; X70817; CAA50148.1; -
CC PIR; S07128; S07128.
CC PIR; S07155; S07155.
CC PIR; S08562; S08562.
CC PIR; S08573; S08573.
CC PIR; S08574; S08574.
CC PIR; A48940; A48940.
CC HSSP; P10845; 3BTA.
CC MEROPS; M27.002; -.
CC InterPro; IPR000395; Bontoxilysin.
CC InterPro; IPR000130; Zn.MTpeptidase.
CC Pfam; PF01742; Peptidase_M27; 1.
CC PRINTS; PR00760; BONTOXILYSIN.
CC ProDom; PD001963; Bontoxilysin; 1.
CC PROSITE; PS00142; ZINC_PROTEASE; 1.
CC Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT INIT_MET 0 0
FT CHAIN 1 440 BOTULINUM NEUROTOXIN B, LIGHT-CHAIN.
FT CHAIN 441 1290 BOTULINUM NEUROTOXIN B, HEAVY-CHAIN.
FT METAL 229 229 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 230 230 BY SIMILARITY.

```



```

FT METAL          233      233      ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID      436      445      INTERCHAIN (PROBABLE).
FT CONFLICT      29        29        T -> M (IN REF. 4).
FT CONFLICT      217      217        R -> G (IN REF. 2).
FT CONFLICT      224      224        A -> S (IN REF. 2).
FT CONFLICT      463      463        S -> R (IN REF. 4).
SQ SEQUENCE      1290 AA; 150670 MW; D21746E2C024DF43 CRC64;

Query Match          51.8%; Score 58; DB 1; Length 1290;
Best Local Similarity 64.3%; Pred. No. 0.15;
Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FNNFTVSFWLRVPK 14
   | : | | | | | : | |
Db 922 FLDFSVSWIRIPK 935

RESULT 6
VP2_AHSV6
ID VP2_AHSV6      STANDARD;      PRT; 1051 AA.
AC O71024;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Outer capsid protein VP2.
GN S2 OR L2.
OS African horse sickness virus 6 (AHSV-6) (African horse sickness virus
OS (serotype 6)).
OC Viruses; dsRNA viruses; Reoviridae; Orbivirus.
OX NCBI_TaxID=86060;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98278331; PubMed=9617769;
RA Williams C.F., Inoue T., Lucas A.-M., Zanotto P., Roy P.;
RT "The complete sequence of four major structural proteins of African
RT horse sickness virus serotype 6: evolutionary relationships within
RT and between the orbiviruses.";
RL Virus Res. 53:53-73(1998).
CC -1- FUNCTION: THE VP2 PROTEIN IS ONE OF THE TWO PROTEINS (WITH VP5)
CC WHICH CONSTITUTE THE VIRUS PARTICLE OUTER CAPSID. IT IS THE
CC MAJOR TARGET OF THE HOST IMMUNOGENIC RESPONSE.
CC -1- SIMILARITY: BELONGS TO THE REOVIRUSES VP2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AF021235; AAC40994.1; .
DR InterPro; IPR001742; Orbi_VP2.
DR Pfam; PF00898; Orbi_VP2; 1.
KW Coat protein.
SQ SEQUENCE 1051 AA; 122326 MW; 2B04DB9E389F4B5F CRC64;

Query Match          50.9%; Score 57; DB 1; Length 1051;
Best Local Similarity 47.6%; Pred. No. 0.18;
Matches 10; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Qy 1 FNNFTVSFWLRVPKVSASHLE 21
   | : | | | | | : | |
Db 636 FSKRFVSFWYRVEKIKTHLE 656

RESULT 7
BXE_CLOBO
ID BXE_CLOBO      STANDARD;      PRT; 1250 AA.
AC Q00436;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)

```

```

DE Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DE (Bontoxilysin E).
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Beluga;
RX MEDLINE=92181428; PubMed=1543481;
RA Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT "Sequences of the botulinum neurotoxin E derived from Clostridium
RT botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT ATCC 43181 and ATCC 43755).";
RL Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92174922; PubMed=1541280;
RA Whelan S.M., Elmore M.J., Bodsworth N.J., Atkinson T., Minton N.P.;
RT "The complete amino acid sequence of the Clostridium botulinum type-E
RT neurotoxin, derived by nucleotide-sequence analysis of the encoding
RT gene.";
RL Eur. J. Biochem. 204:657-667(1992).
RN [3]
RP SEQUENCE OF 1-251 FROM N.A.
RX MEDLINE=90264400; PubMed=2160960;
RA Binz T., Kurazono H., Wille M., Frevert J., Wernars K., Niemann H.;
RT "The complete sequence of botulinum neurotoxin type A and comparison
RT with other clostridial neurotoxins.";
RL J. Biol. Chem. 265:9153-9158(1990).
RN [4]
RP SEQUENCE OF 1-13.
RX MEDLINE=85197963; PubMed=3888113;
RA Schmidt J.J., Sathiyamoorthy V., Dasgupta B.R.;
RT "Partial amino acid sequences of botulinum neurotoxins types B and
RT E.";
RL Arch. Biochem. Biophys. 238:544-548(1985).
RN [5]
RP SEQUENCE OF 419-426.
RX MEDLINE=90344918; PubMed=2116911;
RA Gimenez J.A., Dasgupta B.R.;
RT "Botulinum neurotoxin type E fragmented with endoproteinase Lys-C
RT reveals the site trypsin nicks and homology with tetanus
RT neurotoxin.";
RL Biochimie 72:213-217(1990).
RN [6]
RP IDENTIFICATION OF SUBSTRATE.
RX MEDLINE=94063091; PubMed=8243676;
RA Schiavo G., Santucci A., Dasgupta B.R., Mehta P.P., Jontes J.,
RA Benfenati F., Wilson M.C., Montecucco C.;
RT "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
RT COOH-terminal peptide bonds.";
RL FEBS Lett. 335:99-103(1993).
RN [7]
RP IDENTIFICATION OF SUBSTRATE.
RX MEDLINE=94124495; PubMed=8294407;
RA Binz T., Biasi J., Yamaseki S., Baumeister A., Link E., Suedhof T.C.,
RA Jahn R., Niemann H.;
RT "Proteolysis of SNAP-25 by types E and A botulin neurotoxins.";
RL J. Biol. Chem. 269:1617-1620(1994).
CC -1- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 180-ARG-|-ILE-
CC 181 BOND IN SNAP-25.
CC -1- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC detected action on small molecule substrates.
CC -1- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
CC WHILE THE N-AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL

```

FORMATION AND TOXIN BINDING, RESPECTIVELY.

CC -1- SUBCELLULAR LOCATION: Secreted.

CC -1- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF

CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.

CC -----

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; X62089; CAA43999.1; -

CC EMBL; X62683; CAA44558.1; -

CC PIR; A60027; A60027.

CC PIR; B35294; B35294.

CC PIR; JH0257; JH0257.

CC PIR; S08575; S08575.

CC PIR; S18111; S18111.

CC PIR; S21178; S21178.

CC HSP; P10845; 3BTA.

CC MSERP; M27.002; -

CC InterPro; IPR000395; Bontoxilysin.

CC InterPro; IPR000130; Zn_Mtpeptidase.

CC Pfam; PF01742; Peptidase M27; 1.

CC PRINTS; PR00760; Bontoxilysin.

CC PRODOM; PD001963; Bontoxilysin; 1.

CC PROSITE; PS00142; ZINC_PROTEASE; 1.

CC Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.

CC INIT MET 0

CC CHAIN 1 421

CC CHAIN 422 1250

CC METAL 211 212

CC ACT_SITE 212 215

CC METAL 215 215

CC DISULFID 411 425

CC CONFLICT 176 176

CC CONFLICT 197 197

CC CONFLICT 339 339

CC CONFLICT 772 772

CC CONFLICT 962 962

CC CONFLICT 966 966

CC CONFLICT 1194 1194

CC SEQUENCE 1250 AA; 143712 MW; D9FCE26DDA041EB4 CRC64;

Query Match 50.0%; Score 56; DB 1; Length 1250;

Best Local Similarity 53.8%; Pred. No. 0.32;

Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 FNNFTVSFWLRVP 13

Db 911 YKNFISFWLRIP 923

RESULT 8

BXE_CLOBU

ID_BXE_CLOBU STANDARD; PRT; 1250 AA.

AC P30995;

DT 01-JUL-1993 (Rel. 26, Created)

DT 01-JUL-1993 (Rel. 26, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BONT/E)

DE (Bontoxilysin E).

OS Clostridium butyricum.

OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;

OC Clostridium.

OX NCBI_TaxID=1492;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 43181, and ATCC 43755;

RX MEDLINE=92181428; PubMed=1543481;

RA Poullet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;

RT "Sequences of the botulin neurotoxin E derived from Clostridium

RT botulinum type E (strain Beluga) and Clostridium butyricum (strains

RT ATCC 43181 and ATCC 43755).";

RL Biochem. Biophys. Res. Commun. 183:107-113(1992).

RN [2]

RP SEQUENCE OF 1-251 FROM N.A.

RC STRAIN=BL6340;

RX MEDLINE=91237316; PubMed=2033376;

RA Fujii N., Kimura K., Murakami T., Indoh T., Tsuzuki K.,

RA Yokosawa N., Yashiki T., Oguma K.;

RT "Cloning of a DNA fragment encoding the 5'-terminus of the botulinum

RT type E toxin gene from Clostridium butyricum strain BL6340.";

RL J. Gen. Microbiol. 137:519-525(1991).

RN [3]

RP SEQUENCE OF 1-48.

RC STRAIN=5262;

RA Gimenez J., Foley J., Dasgupta B.R.;

RT "Neurotoxin type E from Clostridium botulinum and C. butyricum;

RT partial sequence and comparison.";

RL FASEB J. 2:A1750-A1750(1988).

CC -1- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER

CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED

CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD

CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT

CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC

CC ENDOPEPTIDASE.

CC -1- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the

CC neurocycytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No

CC detected action on small molecule substrates.

CC -1- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A

CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,

CC WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL

CC FORMATION AND TOXIN BINDING, RESPECTIVELY.

CC -1- SUBCELLULAR LOCATION: Secreted

CC -1- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF

CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.

CC -----

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; X62088; CAA43999.1; -

CC EMBL; X53180; CAA37321.1; -

CC PIR; JH0256; JH0256.

CC PIR; S16145; S16145.

CC HSP; P10845; 3BTA.

CC MSERP; M27.002; -

CC InterPro; IPR000395; Bontoxilysin.

CC InterPro; IPR000130; Zn_Mtpeptidase.

CC Pfam; PF01742; Peptidase M27; 1.

CC PRINTS; PR00760; BONTXILYSIN.

CC PRODOM; PD001963; Bontoxilysin; 1.

CC PROSITE; PS00142; ZINC_PROTEASE; 1.

CC Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.

CC INIT MET 0

CC CHAIN 1 421

CC CHAIN 422 1250

CC METAL 211 211

CC ACT_SITE 212 212

CC METAL 215 215

CC DISULFID 411 425

CC CONFLICT 229 229

CC SEQUENCE 1250 AA; 143265 MW; 8171B5B2C312857 CRC64;

Query Match 50.0%; Score 56; DB 1; Length 1250;

Best Local Similarity 53.8%; Pred. No. 0.32;

Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 FNNFTVSWLRVP 13
DB 911 YKNSISFWVRIP 923

RESULT 9
BXA2 CLOBO
ID BXA2 CLOBO STANDARD; PRT; 1295 AA.
AC Q45894; P77780;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BONT/A)
DE (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-
chain; Botulinum neurotoxin A, heavy-chain].
DE BOTA OR BNA OR ATX.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Kyoto-F;
RX MEDLINE=94143603; PubMed=8310180;
RA Willens A., East A.K., Lawson P.A., Collins M.D.;
RT "Organization and phylogenetic interrelationships of genes encoding
RT components of the botulinum toxin complex in proteolytic Clostridium
RT botulinum types A, B, and F: evidence of chimeric sequences in the
RT gene encoding the nontoxic nonhemagglutinin component.";
RL Int. J. Syst. Bacteriol. 46:1105-1112(1996).
CC -1- FUNCTION: Inhibits acetylcholine release. The botulinum toxin
CC binds with high affinity to peripheral neuronal presynaptic
CC membrane, is then internalized by receptor-mediated endocytosis.
CC The C-terminus of the heavy chain (H) is responsible for the
CC adherence of the toxin to the cell surface while the N-terminus
CC mediates transport of the light chain from the endocytic vesicle
CC to the cytosol. After translocation, the light chain (L)
CC hydrolyzes the 197-Gln-|-Arg-198 bond in SNAP-25, thereby blocking
CC neurotransmitter release. Inhibition of acetylcholine release
CC results in flaccid paralysis, with frequent heart or respiratory
CC failure (By similarity).
CC -1- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC detected action on small molecule substrates.
CC -1- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC heavy chain (H) (By similarity).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- MISCELLANEOUS: There are seven antigenically distinct forms of
CC botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X73423; CAA51824.1; -.
CC DR EMBL; X87974; CAA61234.1; -.
CC DR HSSP; F10845; 3BETA.
CC DR MEROPS; M27.002; -.

QY 1 FNNFTVSWLRVPK 14
DB 937 YENFSTFWIKIPK 950

RESULT 10
VNSS_VNSV STANDARD; PRT; 449 AA.
AC Q01811;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Nonstructural protein NS-S.
GN NS.
OS Impatiens necrotic spot virus (strain NL-07) (INSV).
OC Viruses; ssRNA negative-strand viruses; Bunyaviridae; Tospovirus.
OX NCBI_TaxID=31622;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=922311780; PubMed=1385787;
RA de Haan P., de Avila A.C., Kormelink R., Westerbroek A.,
RA Gielen J.J., Peters D., Goldbach R.;
RT "The nucleotide sequence of the S RNA of Impatiens necrotic spot
RT virus, a novel tospovirus.";
RL FEBS Lett. 306:127-32(1992).
CC -1- FUNCTION: FORMS FILAMENTOUS INCLUSION BODIES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X66972; CAA47382.1; -.
CC DR EMBL; X66972; CAA47382.1; -.
CC DR InterPro; IPR004915; Bunya_NS-S_2.
CC DR Pfam; PF03231; Bunya_NS-S_2; 1.
CC KW Nonstructural protein.
SQ SEQUENCE 449 AA; 51197 MW; C46AC1372B114CA5 CRC64;

Query Match 46.4%; Score 52; DB 1; Length 449;
Best Local Similarity 45.0%; Pred. No. 0.49;
Matches 9; Conservative 5; Mismatches 4; Indels 2; Gaps 1;

QY 2 NN--FTVSWLRVPKVSASH 19
DB 240 NNRPFKISLMMRIPKIMKSN 259

Query Match 46.4%; Score 52; DB 1; Length 449;
Best Local Similarity 45.0%; Pred. No. 0.49;
Matches 9; Conservative 5; Mismatches 4; Indels 2; Gaps 1;

```

```

RESULT 11:
VNSS_TSWV1 STANDARD; PRT; 464 AA.
AC P26002;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Nonstructural protein NS-S.
GN NSS.
OS Tomato spotted wilt virus (strain Brazilian Br-01) (TSWV).
OC Viruses; ssRNA negative-strand viruses; Bunyaviridae; Tospovirus.
OX NCBI_TaxID=36413;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90264829; PubMed=1693160;
RA de Haan P., Wagmakers L., Peters D., Goldbach R.;
RT "The S RNA segment of tomato spotted wilt virus has an ambisense
character.";
RL J. Gen. Virol. 71:1001-1007(1990).
CC -!- FUNCTION: FORMS FILAMENTOUS INCLUSION BODIES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D00645; BAA00540.1; -.
DR PIR; JQ0547; MNVUWC.
DR InterPro; IPR004915; Bunya NS-S_2.
DR Pfam; PF03231; Bunya NS-S_2; 1.
KW Nonstructural protein.
SQ SEQUENCE 464 AA; 52448 MW; AE5519179F9BF377 CRC64;

Query Match 46.4%; Score 52; DB 1; Length 464;
Best Local Similarity 64.3%; Pred. No. 0.51;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 NNFTVSFWLRVPKV 15
DB 239 SHFKLSLWLRVPKV 252

RESULT 12:
VNSS_TSWV1 STANDARD; PRT; 467 AA.
AC P26003;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Nonstructural protein NS-S.
GN NSS.
OS Tomato spotted wilt virus (strain Bulgarian L3) (TSWV).
OC Viruses; ssRNA negative-strand viruses; Bunyaviridae; Tospovirus.
OX NCBI_TaxID=36415;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91132150; PubMed=1993884;
RA Maiss E., Ivanova L., Breyel E., Adam G.;
RT "Cloning and sequencing of the S RNA from a Bulgarian isolate of
tomato spotted wilt virus.";
RL J. Gen. Virol. 72:461-464(1991).
CC -!- FUNCTION: FORMS FILAMENTOUS INCLUSION BODIES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D00645; BAA00540.1; -.
DR PIR; JQ0547; MNVUWC.
DR InterPro; IPR004915; Bunya NS-S_2.
DR Pfam; PF03231; Bunya NS-S_2; 1.
KW Nonstructural protein.
SQ SEQUENCE 464 AA; 52448 MW; AE5519179F9BF377 CRC64;

Query Match 46.4%; Score 52; DB 1; Length 464;
Best Local Similarity 64.3%; Pred. No. 0.51;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 NNFTVSFWLRVPKV 15
DB 239 SHFKLSLWLRVPKV 252

RESULT 13:
BXCN_CLOBO STANDARD; PRT; 1196 AA.
AC P46081;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Botulinum neurotoxin type C1, nontoxic component.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Type C Stockholm / C-ST;
RX MEDLINE=92231894; PubMed=1567404;
RA Tsuzuki K., Kimura K., Fujii N., Yokosawa N., Oguma K.;
RT "The complete nucleotide sequence of the gene coding for the
RT nontoxic-nonhemagglutinin component of Clostridium botulinum type C
RT progenitor toxin.";
RL Biochem. Biophys. Res. Commun. 183:1273-1279(1992).
CC -!- FUNCTION: THE NONTXIC COMPONENT IS NECESSARY TO MAINTAIN
CC TOXICITY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X62389; CAA44262.1; -.
DR InterPro; IPR000395; Bontoxilysin.
DR Pfam; PF01742; Peptidase M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
KW Neurotoxin.
SQ SEQUENCE 1196 AA; 138740 MW; 4BD5956274D7F9C3 CRC64;

Query Match 45.5%; Score 51; DB 1; Length 1196;
Best Local Similarity 80.0%; Pred. No. 2.1;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 NNFTVSFWLR 11
DB 899 NNFTICFWLR 908

RESULT 14:
BXCI_CLOBO STANDARD; PRT; 1290 AA.
AC P18640;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)

```


Search completed: July 22, 2003, 08:06:40
Job time : 1.55306 secs

RA MEDLINE=94230352; PubMed=8175689;
RA Yamasaki S., Baumeister A., Binz T., Blasi J., Link E., Cornille F.,
RA Roques B., Fyke E.M., Suedhof T.C., Jahn R., Niemann H.;
RT "Cleavage of members of the synaptobrevin/VAMP family by types D and
RT F botulinum neurotoxins and tetanus toxin.";
RL J. Biol. Chem. 269:12764-12772(1994).
CC -1- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE THAT CLEAVES THE 60-LYS-|-LEU-61 BOND OF
CC SYNAPTOSOMES-1 AND -2.
CC -1- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC neurocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC detected action on small molecule substrates.
CC -1- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
CC WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL
CC FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC -1- MISCELLANEOUS: BOTULINUM TYPE D NEUROTOXIN IS SYNTHESIZED BY D
CC STRAIN OF CLOSTRIDIUM BOTULINUM WHICH CARRY THE APPROPRIATE
CC BACTERIOPHAGE.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X54254; CAA38175.1; -;
DR EMBL; S49407; AAB24244.1; -;
DR PIR; S11455; S11455.
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -;
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR000130; Zn_Mtpeptidase.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
KW CHAIN 1 442 BOTULINUM NEUROTOXIN D, LIGHT-CHAIN.
FT CHAIN 443 1276 BOTULINUM NEUROTOXIN D, HEAVY-CHAIN.
FT METAL 229 229 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT SITE 230 230 BY SIMILARITY.
FT METAL 233 233 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 437 450 INTERCHAIN (PROBABLE).
FT VARIANT 15 16 ND -> PV (IN STRAIN D-SA).
FT VARIANT 17 18 ND -> LQ (IN STRAIN D-1873).
FT VARIANT 452 452 K -> Q (IN STRAIN D-SA).
FT VARIANT 457 457 R -> T (IN STRAIN D-SA).
FT VARIANT 457 457 R -> F (IN STRAIN D-1873).
FT VARIANT 462 462 A -> D (IN STRAIN D-1873).
FT VARIANT 489 489 K -> N (IN STRAIN CB16).
FT VARIANT 644 644 N -> K (IN STRAIN CB16).
FT VARIANT 1122 1122 Q -> R (IN STRAIN CB16).
SQ SEQUENCE 1276 AA; 146871 MW; C1EC50F46C8233E2 CRC64;

Query Match 42.4%; Score 47.5; DB 1; Length 1276;
Best Local Similarity 40.9%; Pred.No. 8.3;
Matches 9; Conservative 7; Mismatches 5; Indels 1; Gaps 1;

QY 1 FNNFTVSFWLRVPEK-VSASHLE 21

Db 928 YENSSVSFWIKSKDLTNSHNE 949

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:05:49 ; Search time 2.11464 Seconds
(without alignments)
2046.206 Million cell updates/sec

Title: US-09-806-703A-14
Perfect score: 112
Sequence: 1 FNNFTVFWLRVPKVSASHLE 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPREMBL 21:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	112	100.0	451	2	Q9LA13
2	112	100.0	1310	2	Q93N27
3	62	55.4	1268	2	Q45851
4	61	54.5	1278	2	Q57236
5	58	51.8	361	2	Q45846
6	58	51.8	361	2	Q45848
7	58	51.8	441	2	Q9X708
8	58	51.8	1291	2	Q93G71
9	58	51.8	1291	2	Q933K0
10	58	51.8	1291	2	Q9ZAJ8
11	58	51.8	1291	2	Q08077
12	56	50.0	367	2	Q45861
13	56	50.0	367	2	Q45862
14	56	50.0	1251	2	Q9K395
15	56	50.0	1255	2	Q9FAR6
16	55	49.1	1280	2	Q9ZAJ5

17	52	46.4	467	12	Q37367	O37367 tomato spot
18	52	46.4	467	12	Q37369	O37369 tomato spot
19	51	45.5	467	12	Q88900	Q88900 tospovirus.
20	51	45.5	1196	2	Q9LBS8	Q9LBS8 clostridium
21	51	45.5	1196	2	Q9LBR2	Q9LBR2 clostridium
22	51	45.5	1196	2	Q53550	Q53550 clostridium
23	51	45.5	1196	2	Q45916	Q45916 clostridium
24	51	45.5	1196	2	Q93HT4	Q93HT4 clostridium
25	51	45.5	1196	9	Q9ZX77	Q9ZX77 clostridium
26	51	45.5	1196	9	Q38197	Q38197 clostridium
27	50	44.6	503	10	Q948E8	Q948E8 oryza sativ
28	49	43.8	276	5	Q97ZK5	Q97ZK5 caenorhabdi
29	48	42.9	504	10	Q9WIN3	Q9WIN3 arabidopsis
30	48	42.9	1285	2	Q9LBR1	Q9LBR1 clostridium
31	48	42.9	1285	2	Q45967	Q45967 clostridium
32	48	42.9	1291	2	Q93HT3	Q93HT3 clostridium
33	47.5	42.4	1275	12	Q9QTG7	Q9QTG7 clostridium
34	47.5	42.4	1280	2	Q9LBS7	Q9LBS7 clostridium
35	47.5	42.4	1280	2	Q45849	Q45849 clostridium
36	47	42.0	345	5	O62165	O62165 caenorhabdi
37	47	42.0	760	2	Q9S429	Q9S429 clostridium
38	46	41.1	278	2	Q93JMO	Q93JMO streptomyce
39	46	41.1	322	5	P91551	P91551 caenorhabdi
40	46	41.1	442	6	Q07452	Q07452 oviz aries
41	46	41.1	496	3	Q9UUI6	Q9UUI6 schizosacch
42	46	41.1	558	5	O76933	O76933 drosophila
43	46	41.1	558	5	Q9V3N9	Q9V3N9 drosophila
44	46	41.1	753	10	Q9C7W5	Q9C7W5 arabidopsis
45	46	41.1	996	10	Q9C925	Q9C925 arabidopsis

ALIGNMENTS

RESULT 1

Q9LA13 ID Q9LA13 PRELIMINARY; PRT; 451 AA.
AC Q9LA13;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Tetanus toxin (fragment).
OS Clostridium tetani.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1513;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=20886;
RA He H.J., Shi H.J., Fe Z.Y., Yuan Q.S., Wu X.F.;
RT "Fragment C of Tetanus Toxin."
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF154828; AAF73267.1; -
DR HSSP; P04958; IABD.
DR InterPro; IPR001064; CRYSTALLIN.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
FT NON TER 1
SQ SEQUENCE 451 AA; 51823 MW; 69A8C5F030E6CD8E CRC64;

Query Match 100.0%; Score 112; DB 2; Length 451;
Best Local Similarity 100.0%; Pred. No. 1.7e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FNNFTVFWLRVPKVSASHLE 21
Db 83 FNNFTVFWLRVPKVSASHLE 103

RESULT 2

Q93N27 ID Q93N27 PRELIMINARY; PRT; 1310 AA.
AC Q93N27;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)

01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 Tetanus toxin (Fragment).
 Clostridium tetani.
 Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
 Clostridiales; Clostridiaceae; Clostridium.
 NCBI_TaxID=1513;
 [1]
 SEQUENCE FROM N.A.
 Shumin Z., Dianliang L.;
 "Cloning and sequence analysis of tetanus toxin gene.";
 Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 EMBL; AF389424; AAK72964.2;
 InterPro; IPR000395; Bontoxilysin.
 InterPro; IPR001064; Crystallin.
 InterPro; IPR000130; Zn_MTPetdse.
 Pfam; PF01742; Peptidase M27; 1.
 ProDom; PD001963; Bontoxilysin; 1.
 PROSITE; PS00225; CRYSTALLIN BETAGAMMA; UNKNOWN_1.
 PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
 NON_TER 1
 NON_TER 1310 1310
 SQ SEQUENCE 1310 AA; 150316 MW; 9EADDC914418E450 CRC64;

Query Match 100.0%; Score 112; DB 2; Length 1310;
 Best Local Similarity 100.0%; Pred. No. 5.2e-10;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FNNFTVSFWLRVPKVSASHLE 21
 |||||
 Db 948 FNNFTVSFWLRVPKVSASHLE 968

RESULT 3

Q45851 ID Q45851 PRELIMINARY; PRT; 1268 AA.
 AC Q45851;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Neurotoxin type F.
 GN BONT / F.
 OS Clostridium baratii.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
 OC Clostridiales; Clostridiaceae; Clostridium.
 NCBI_TaxID=1561;
 [1]
 SEQUENCE FROM N.A.
 MEDLINE=9325228; PubMed=8486245;
 Thompson D.E., Hutson R.A., East A.K., Allaway D., Collins M.D.,
 Richardson P.T.;
 "Nucleotide sequence of the gene coding for Clostridium baratii type F
 neurotoxin: Comparison with other clostridial neurotoxins";
 FEMS Microbiol. Lett. 108:175-182 (1993).
 EMBL; X68262; CAA48329.1;
 HSSP; P10845; 3BTA.
 MEROPS; M27.002;
 InterPro; IPR000395; Bontoxilysin.
 Pfam; PF01742; Peptidase M27; 1.
 PRINTS; PR00760; BONTXILYSIN.
 ProDom; PD001963; Bontoxilysin; 1.
 PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
 SQ SEQUENCE 1268 AA; 145513 MW; 963040091AC15ED2 CRC64;

Query Match 55.4%; Score 62; DB 2; Length 1268;
 Best Local Similarity 64.3%; Pred. No. 0.13;
 Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FNNFTVSFWLRVPK 14
 : |||:|:|:|:|
 Db 922 YQNFSISFWVRPK 935

RESULT 4

Q57236 ID Q57236 PRELIMINARY; PRT; 1278 AA.
 AC Q57236; Q45863;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE Botulinum neurotoxin type F (BONT/F protein).
 GN BONT/F.
 OS Clostridium botulinum.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
 OC Clostridiales; Clostridiaceae; Clostridium.
 NCBI_TaxID=1491;
 [1]
 SEQUENCE FROM N.A.
 RC STRAIN=NCTC 10281;
 RA Hutson R.A., Collins M.D.;
 Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 SEQUENCE FROM N.A.
 RA Elmore M.J., Bodsworth N.J., Whelan S.M., Minton N.P.;
 Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
 RN [3]
 SEQUENCE OF 635-1000 FROM N.A.
 RC STRAIN=NCTC 1028;
 RX MEDLINE=94013172; PubMed=8408542;
 RA Campbell K., East A.K., Collins M.D.;
 "Gene probes for identification of the botulin neurotoxin gene and
 specific identification of neurotoxin types B, E, and F";
 J. Clin. Microbiol. 31:2255-2262 (1993).
 RN [4]
 SEQUENCE OF 1-27 FROM N.A.
 RC STRAIN=LANGELAND;
 RX MEDLINE=98404102; PubMed=9732534;
 RA East A.K., Bhandari M., Hielm S., Collins M.D.;
 "Analysis of the botulinum neurotoxin type F gene clusters in
 proteolytic and nonproteolytic Clostridium botulinum and Clostridium
 baratii";
 Curr. Microbiol. 37:262-268 (1998).
 DR EMBL; X81714; CAA57358.1;
 DR EMBL; L35496; AAA23210.1;
 DR EMBL; X70821; CAA50152.1;
 DR EMBL; X99064; CAA67512.1;
 DR HSSP; P10845; 3BTA.
 DR MEROPS; M27.002;
 DR InterPro; IPR000395; Bontoxilysin.
 DR InterPro; IPR000130; Zn_MTPetdse.
 DR Pfam; PF01742; Peptidase M27; 1.
 DR PRINTS; PR00760; BONTXILYSIN.
 DR ProDom; PD001963; Bontoxilysin; 1.
 DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
 KW Neurotoxin.
 SQ SEQUENCE 1278 AA; 147073 MW; A1BE1318431D6918 CRC64;

Query Match 54.5%; Score 61; DB 2; Length 1278;
 Best Local Similarity 57.1%; Pred. No. 0.19;
 Matches 8; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FNNFTVSFWLRVPK 14
 : |||:|:|:|:|
 Db 931 YQNFSISFWVRPK 944

RESULT 5

Q45846 ID Q45846 PRELIMINARY; PRT; 361 AA.
 AC Q45846;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 DE Botulinum neurotoxin type B (Fragment).
 GN BONT/B.

```

OS Clostridium botulinum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TYPE B;
RX MEDLINE=94013372; PubMed=8408542;
RA Campbell K., East A.K., Collins M.D.;
RT "Gene probes for identification of the botulin neurotoxin gene and
RT specific identification of neurotoxin types B, E, and F.";
RL J. Clin. Microbiol. 31:2255-2262(1993).
DR EMBL; X70814; CAA50145.1; -.
DR HSSP; P10845; 3BTA.
KW Neurotoxin.
FT NON_TER 1 1
FT NON_TER 361 361
SQ SEQUENCE 361 AA; 42175 MW; 533EA98735CD98E1 CRC64;

Query Match 51.8%; Score 58; DB 2; Length 361;
Best Local Similarity 64.3%; Pred. No. 0.16;
Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 FNNFTVSFWLRVPK 14
Db 290 FLDFSVSWIRIPK 303

RESULT 6
Q45848 PRELIMINARY; PRT; 361 AA.
AC Q45848;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE Botulinum neurotoxin type B (Fragment).
GN BONT/B.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TYPE B;
RX MEDLINE=94013372; PubMed=8408542;
RA Campbell K., East A.K., Collins M.D.;
RT "Gene probes for identification of the botulin neurotoxin gene and
RT specific identification of neurotoxin types B, E, and F.";
RL J. Clin. Microbiol. 31:2255-2262(1993).
DR EMBL; X70819; CAA50150.1; -.
DR HSSP; P10845; 3BTA.
KW Neurotoxin.
FT NON_TER 1 1
FT NON_TER 361 361
SQ SEQUENCE 361 AA; 42131 MW; A2E0FFFC81F9533D CRC64;

Query Match 51.8%; Score 58; DB 2; Length 361;
Best Local Similarity 64.3%; Pred. No. 0.16;
Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 FNNFTVSFWLRVPK 14
Db 290 FLDFSVSWIRIPK 303

RESULT 7
Q9X708 PRELIMINARY; PRT; 441 AA.
AC Q9X708;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Botulinum neurotoxin type B (Fragment).

GN BONT/B.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9343691; PubMed=10413679;
RA Lalli G., Herreros J., Osborne S.L., Montecucco C., Rossetto O.,
RA Schiavo G.;
RT "Functional characterisation of tetanus and botulinum neurotoxins
RT binding domains.";
RL J. Cell Sci. 112:2715-2724(1999).
DR EMBL; AJ242628; CAB43706.1; -.
DR HSSP; P10845; 3BTA.
KW Neurotoxin.
FT NON_TER 1 1
FT NON_TER 441 441
SQ SEQUENCE 441 AA; 52772 MW; 721D0B468E8C95A4 CRC64;

Query Match 51.8%; Score 58; DB 2; Length 441;
Best Local Similarity 64.3%; Pred. No. 0.2;
Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 FNNFTVSFWLRVPK 14
Db 81 FLDFSVSWIRIPK 94

RESULT 8
Q93G71 PRELIMINARY; PRT; 1291 AA.
AC Q93G71;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Neurotoxin type B.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1436;
RA Kirma N., Ferreira J.L., Baumstark B.R.;
RT "Characterization of six type A strains of Clostridium botulinum that
RT contain type B toxin gene sequences.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF295926; AAK97132.1; -.
DR InterPro; IPR000130; Zn.MTpeptdse.
DR Pfam; PF01742; Peptidase_M27; 1.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN 1.
SQ SEQUENCE 1291 AA; 150824 MW; D7CA07BAE2EB8CD2 CRC64;

Query Match 51.8%; Score 58; DB 2; Length 1291;
Best Local Similarity 64.3%; Pred. No. 0.6;
Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 FNNFTVSFWLRVPK 14
Db 923 FLDFSVSWIRIPK 936

RESULT 9
Q933K0 PRELIMINARY; PRT; 1291 AA.
AC Q933K0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Type B cryptic neurotoxin.

```



```

RP SEQUENCE FROM N.A.
RC STRAIN=TYPE E;
RX MEDLINE=94013372; PubMed=8408542;
RA Campbell K., East A.K., Collins M.D.;
RT "Gene probes for identification of the botulinum neurotoxin gene and
RT specific identification of neurotoxin types B, E, and F.";
RL J. Clin. Microbiol. 31:2255-2262(1993).
DR EMBL; X70818; CAA50149.1; -.
DR HSSP; P10845; 3BTA.
KW Neurotoxin.
FT NON_TER 1 1
FT NON_TER 367 367
SQ SEQUENCE 367 AA; 42902 MW; 346A610C2FF70262 CRC64;

Query Match 50.0%; Score 56; DB 2; Length 367;
Best Local Similarity 53.8%; Pred. No. 0.36;
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 FNNFTVSFWLRVP 13
Db 297 YKNFSISFWVRIP 309

RESULT 13
Q45862 PRELIMINARY; PRT; 367 AA.
AC Q45862;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE Botulinum neurotoxin type E (Fragment).
GN BONT/E.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1491;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=TYPE E;
RX MEDLINE=94013372; PubMed=8408542;
RA Campbell K., East A.K., Collins M.D.;
RT "Gene probes for identification of the botulinum neurotoxin gene and
RT specific identification of neurotoxin types B, E, and F.";
RL J. Clin. Microbiol. 31:2255-2262(1993).
DR EMBL; X70815; CAA50146.1; -.
DR HSSP; P10845; 3BTA.
KW Neurotoxin.
FT NON_TER 1 1
FT NON_TER 367 367
SQ SEQUENCE 367 AA; 42854 MW; 0810595B3A865570 CRC64;

Query Match 50.0%; Score 56; DB 2; Length 367;
Best Local Similarity 53.8%; Pred. No. 0.36;
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 FNNFTVSFWLRVP 13
Db 297 YKNFSISFWVRIP 309

RESULT 14
Q9K395 PRELIMINARY; PRT; 1251 AA.
AC Q9K395;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Type E botulinum toxin.
GN BONT/E.
OS Clostridium butyricum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1492;

RN
RP SEQUENCE FROM N.A.
RC STRAIN=LCL 095;
RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA Karasawa T.;
RT "C. butyricum (LCL 095) gene for type E botulinum toxin.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LCL 155;
RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Gyobu Y., Yamakawa K.,
RA Kato H., Nakamura S., Karasawa T.;
RT "C. butyricum (LCL 155) gene for type E botulinum toxin.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=KZ 1899;
RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA Karasawa T.;
RT "C. butyricum (KZ 1899) gene for type E botulinum toxin.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=KZ 1897;
RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA Karasawa T.;
RT "C. butyricum (KZ 1897) gene for type E botulinum toxin.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=KZ 1898;
RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA Karasawa T.;
RT "C. butyricum (KZ 1898) gene for type E botulinum toxin.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=KZ 1886;
RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA Karasawa T.;
RT "C. butyricum (KZ 1886) gene for type E botulinum toxin.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE FROM N.A.
RC STRAIN=KZ 1887;
RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA Karasawa T.;
RT "C. butyricum (KZ 1887) gene for type E botulinum toxin.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [8]
RP SEQUENCE FROM N.A.
RC STRAIN=KZ 1889;
RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA Karasawa T.;
RT "C. butyricum (KZ 1889) gene for type E botulinum toxin.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [9]
RP SEQUENCE FROM N.A.
RC STRAIN=KZ 1890;
RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA Karasawa T.;
RT "C. butyricum (KZ 1890) gene for type E botulinum toxin.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [10]
RP SEQUENCE FROM N.A.
RC STRAIN=KZ 1891;
RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA Karasawa T.;
RT "C. butyricum (KZ 1891) gene for type E botulinum toxin.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [11]
RP SEQUENCE FROM N.A.
RC STRAIN=LCL 063;

```

RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA Karasawa T.,
RT "C. butyricum (ICL 063) gene for type E botulinum toxin."
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR ENBL; AB037714; BAB03522.1; -
DR ENBL; AB037704; BAB03512.1; -
DR ENBL; AB037705; BAB03513.1; -
DR ENBL; AB037706; BAB03514.1; -
DR ENBL; AB037707; BAB03515.1; -
DR ENBL; AB037708; BAB03516.1; -
DR ENBL; AB037709; BAB03517.1; -
DR ENBL; AB037710; BAB03518.1; -
DR ENBL; AB037711; BAB03519.1; -
DR ENBL; AB037712; BAB03520.1; -
DR ENBL; AB037713; BAB03521.1; -
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR000130; Zn_MTPeptidase.
DR Pfam; PF01742; Peptidase M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN 1.
SQ SEQUENCE 1251 AA; 143751 MW; 2021F4E427070296 CRC64;

Query Match 50.0%; Score 56; DB 2; Length 1251;
Best Local Similarity 53.8%; Pred. No. 1.3;
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 FNNFTVSFWLRVP 13
:|:|:|:|:
Db 912 YKNFSISFWVRIP 924

RESULT 15

ID Q9FAR6 PRELIMINARY; PRT; 1255 AA.
AC Q9FAR6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Type E botulinum toxin.
GN BONT/E.
OS Clostridium butyricum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1492;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BL 6340/ATCC 43755/BL 5520/KZ 147;
RX MEDLINE=20509829; PubMed=11055954;
RA Wang X., Maegawa T., Karasawa T., Kozaki S., Tsukamoto K., Gyobu Y.,
RA Yamakawa K., Oguma K., Sakaguchi Y., Nakamura S.,
RT "Genetic Analysis of Type E Botulinum Toxin-Producing Clostridium
RT butyricum Strains."
RL Appl. Environ. Microbiol. 66:4992-4997(2000).
DR ENBL; AB039264; BAB12249.1; -
DR HSSP; P10845; 3BTA.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR000130; Zn_MTPeptidase.
DR Pfam; PF01742; Peptidase M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN 1.
SQ SEQUENCE 1255 AA; 143918 MW; 1B557B9D85CD8E4D CRC64;

Query Match 50.0%; Score 56; DB 2; Length 1255;
Best Local Similarity 53.8%; Pred. No. 1.3;
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 FNNFTVSFWLRVP 13
:|:|:|:|:
Db 915 YKNFSISFWVRIP 927

Search completed: July 22, 2003, 08:11:31
Job time : 4.11464 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:05:49 ; Search time 1.05732 Seconds
(without alignments)
1909.378 Million cell updates/sec

Title: US-09-806-703A-14
Perfect score: 112
Sequence: 1 FNNFTVSFWLRVPKVSASHLE 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 73.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	112	100.0	1315	1 BTCLTN	tentoxilysin (EC 3
2	62	55.4	1268	2 S33411	botulinum neurotox
3	61	54.5	366	2 S48110	neurotoxin type F
4	61	54.5	369	2 S48109	neurotoxin type F
5	61	54.5	1274	2 I40813	neurotoxin - Clost
6	61	54.5	1297	2 S39791	neurotoxin - Clost
7	59	52.7	1296	1 BTCLAB	botulinum neurotox
8	58	51.8	1291	1 A48940	botulinum neurotox
9	58	51.8	1291	2 I40631	non-proteolytic bo
10	56	50.0	367	2 S48106	neurotoxin type E
11	56	50.0	1251	2 JH0256	botulinum neurotox
12	56	50.0	1252	2 S21178	botulinum neurotox
13	56	50.0	1296	2 I40645	botulinum neurotox
14	52	46.4	449	2 S23158	nucleocapsid prote
15	52	46.4	464	1 MNVJWC	nonstructural prot
16	52	46.4	467	1 MNVJWC	nonstructural prot
17	51	45.5	1196	2 J01467	toxin, nontoxic co
18	51	45.5	1196	2 S46430	botulinum neurotox
19	49	43.8	276	2 T33493	hypothetical prote
20	48	42.9	504	2 T47446	botulinum neurotox
21	48	42.9	1285	2 S70582	botulinum neurotox
22	48	42.9	1291	2 S46431	botulinum neurotox
23	48	42.9	1276	2 S11455	botulinum neurotox
24	47.5	42.4	1276	2 S11455	botulinum neurotox
25	47	42.0	359	2 F87937	protein F1486.6 [i
26	47	42.0	385	2 T20879	hypothetical prote
27	47	42.0	439	2 B37837	probable alpha-amy
28	46	41.1	322	2 T25966	hypothetical prote
29	46	41.1	442	2 I47074	gene CD5 protein -

30	46	41.1	496	2 T38197	probable myb-like
31	46	41.1	753	2 C96668	unknown protein F1
32	46	41.1	1162	2 A47708	progenitor toxin n
33	46	41.1	1162	2 I40817	botulinum toxin no
34	45	40.2	528	2 T41362	hypothetical prote
35	45	40.2	886	2 T39081	hypothetical prote
36	44	39.3	209	2 A82470	conserved hypochet
37	44	39.3	404	2 AD0574	enterochelin ester
38	44	39.3	591	2 S04401	spheroide monoox
39	43	38.4	152	2 D64943	probable membrane
40	43	38.4	152	2 B99945	hypothetical prote
41	43	38.4	152	2 E85793	hypothetical prote
42	43	38.4	381	2 A90711	hypothetical prote
43	43	38.4	381	2 E85561	hypothetical prote
44	43	38.4	514	2 T20858	hypothetical prote
45	43	38.4	781	2 F83884	hypothetical prote

ALIGNMENTS

RESULT 1

BTCLTN
tentoxilysin (EC 3.4.24.68) precursor - Clostridium tetani
N;Alternate names: tetanus neurotoxin
C;Species: Clostridium tetani
C;Date: 31-Mar-1988 #sequence revision 31-Mar-1988 #text change 03-Jun-2002
C;Accession: A25689; A25757; A25194; B25194; A60759; S69348; S09364
R;Eisel, U.; Jarausch, W.; Goretzki, K.; Henschen, A.; Engels, J.; Weller, U.; Hudel, M.,
EMBO J. 5, 2495-2502, 1986
A;Title: Tetanus toxin: primary structure, expression in E. coli, and homology with botu
A;Reference number: A25689; MUID:87053814; PMID:3536478
A;Accession: A25689
A;Molecule type: DNA
A;Residues: 1-1315 <EIS>
A;Cross-references: GB:X04436; NID:940769; PIDN:CAA28033.1; PID:940770
R;Fairweather, N.F.; Lyness, V.A.
Nucleic Acids Res. 14, 7809-7812, 1986
A;Title: The complete nucleotide sequence of tetanus toxin.
A;Reference number: A25757; MUID:87040747; PMID:3774547
A;Accession: A25757
A;Molecule type: DNA
A;Residues: 1-1315 <FAI>
A;Cross-references: GB:X06214; NID:940773; PIDN:CAA29564.1; PID:940774
R;Fairweather, N.F.; Lyness, V.A.; Pickard, D.J.; Allen, G.; Thomson, R.O.
J. Bacteriol. 185, 21-27, 1986
A;Title: Cloning, nucleotide sequencing, and expression of tetanus toxin fragment C in E
A;Reference number: A25194; MUID:86085672; PMID:3510187
A;Accession: A25194
A;Molecule type: DNA
A;Residues: 743-1315 <FA2>
A;Cross-references: GB:M12739; NID:9144920; PIDN:AAA23282.1; PID:9144921
A;Accession: B25194
A;Molecule type: protein
A;Residues: 865-894 <FA3>
R;Matsuda, M.; Lei, D.L.; Sugimoto, N.; Ozutsumi, K.; Okabe, T.
Infect. Immun. 57, 3588-3593, 1989
A;Title: Isolation, purification, and characterization of fragment B, the NH-2-terminal t
A;Reference number: A60759; MUID:90035436; PMID:2478476
A;Accession: A60759
A;Molecule type: protein
A;Residues: 461-475 <MAT>
R;Demotz, S.; Lanzavecchia, L.; Eisel, U.; Niemann, H.; Widmann, C.; Corradin, G.
J. Immunol. 142, 394-402, 1989
A;Title: Delineation of several DR-restricted tetanus toxin T cell epitopes.
A;Reference number: JS0098; MUID:89093918; PMID:2463305
A;Contents: annotation; epitope region
R;Schlavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta, B.R.
Nature 359, 832-835, 1992
A;Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolyti
A;Reference number: S27125; MUID:93063293; PMID:1331807
A;Contents: annotation

R;de Filippis, V.; Vangelista, L.; Schiavo, G.; Tonello, F.; Montecucco, C.
Eur. J. Biochem. 229, 61-69, 1995
A;Title: Structural studies on the zinc-endopeptidase light chain of tetanus neurotoxin.
A;Reference number: S69348; MUID:95262688; PMID:7744050
A;Accession: S69348
A;Molecule type: protein
A;Residues: 2-31 <DEF>
C;Comment: The source of this protein was an extrachromosomal plasmid.
C;Comment: The precursor is cleaved by endogenous proteinase activity to form light (fragment A) and heavy (fragment B) chains. The amino end of the heavy chain (fragment B) forms ion channels in a lipid bilayer. Fragment C binds to ganglionic presynaptic neurons. It inhibits neurotransmitter release by proteolytic cleavage of synaptobrevin.
A;Description: blocks neuroexocytosis via hydrolysis of a Gln-Phe peptide bond in synaptobrevin.
C;Superfamily: tetanus toxin
C;Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc
F;2-457/Product: tentoxylisin light chain (fragment A) #status predicted <TTL>
F;461-1315/Product: tentoxylisin heavy chain (fragment B.C) #status experimental <TTH>
F;461-864/Domain: channel forming (fragment B) #status predicted <TXB>
F;865-1315/Domain: ganglioside binding (fragment C) #status predicted <TXC>
F;233,237/Binding site: zinc (His) #status predicted
F;234/Active site: Glu #status predicted

Query Match 100.0%; Score 112; DB 1; Length 1315;
Best Local Similarity 100.0%; Pred. No. 3.7e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNNFTVSFWLRVPKVSASHLE 21
Db 947 FNNFTVSFWLRVPKVSASHLE 967

RESULT 2

S33411
botulinum neurotoxin type F - Clostridium baratii
C;Species: Clostridium baratii
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 16-Jul-1999
C;Accession: S33411; S31860
R;Thompson, D.E.; Hutson, R.A.; East, A.K.; Allaway, D.; Collins, M.D.; Richardson, P.T.
FEMS Microbiol. Lett. 108, 175-182, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium baratii type F neurotoxin
A;Reference number: S33411; MUID:93252228; PMID:8486245
A;Accession: S33411
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1268 <THO>
A;Cross-references: EMBL:X68262; NID:G49138; PIDN:CAA48329.1; PID:G49139
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 55.4%; Score 62; DB 2; Length 1268;
Best Local Similarity 64.3%; Pred. No. 0.073;
Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 FNNFTVSFWLRVPK 14
Db 922 YQNFSISFWVRPK 935

RESULT 3

S48110
neurotoxin type F - Clostridium botulinum (fragment)
C;Species: Clostridium botulinum
C;Date: 14-Jul-1995 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C;Accession: S48110
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinum neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48110
A;Status: preliminary; translation not shown
A;Molecule type: DNA

A;Residues: 1-366 <CAM>
A;Cross-references: EMBL:X70821; NID:G407792; PIDN:CAA50152.1; PID:G407793
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 54.5%; Score 61; DB 2; Length 366;
Best Local Similarity 57.1%; Pred. No. 0.028;
Matches 8; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 FNNFTVSFWLRVPK 14
Db 297 YQNFSISFWVRPK 310

RESULT 4

S48109
neurotoxin type F - Clostridium botulinum (fragment)
C;Species: Clostridium botulinum
C;Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 16-Jul-1999
C;Accession: S48109
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinum neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48109
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-369 <CAM>
A;Cross-references: EMBL:X70820; NID:G407790; PIDN:CAA50151.1; PID:G407791
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
C;Superfamily: tetanus toxin

Query Match 54.5%; Score 61; DB 2; Length 369;
Best Local Similarity 57.1%; Pred. No. 0.029;
Matches 8; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 FNNFTVSFWLRVPK 14
Db 297 YQNFSISFWVRPK 310

RESULT 5

I40813
neurotoxin type F - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40813; S48108
R;East, A.K.; Richardson, P.T.; Allaway, D.; Collins, M.D.; Roberts, T.A.; Thompson, D.E.
FEMS Microbiol. Lett. 96, 225-230, 1992
A;Title: Sequence of the gene encoding type F neurotoxin of Clostridium botulinum.
A;Reference number: I40644
A;Accession: I40813
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1274 <RES>
A;Cross-references: GB:M92906; NID:G144866; PIDN:AAA23263.1; PID:G144867
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinum neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372; PMID:8408542
A;Accession: S48108
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 634-1002 <CAM>
A;Cross-references: EMBL:X70816; NID:G407788; PIDN:CAA50147.1; PID:G407789
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 54.5%; Score 61; DB 2; Length 1274;
Best Local Similarity 57.1%; Pred. No. 0.11;
Matches 8; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 FNNFTVSFWLRVPK 14

A;Cross-references: GB:M

A;Accession: S48105

;Refere

;Refere

Search completed: July 22, 2003, 08:12:44
Job time : 2.05732 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:05:44 ; Search time 91.3788 Seconds
(without alignments)
1830.070 Million cell updates/sec

Title: US-09-806-703A-4
Perfect score: 6812
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTFAENPEYGLDVPV 1255

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6812	100.0	1255	21 AAY92620	Human heregulin 2
2	6812	100.0	1255	22 AAE12130	Human tyrosine kin
3	6812	100.0	1255	22 AAB60167	HER2 transgene pla
4	6812	100.0	1255	23 AAU74545	Human HER2 (erbB2)
5	6806	99.9	1255	17 AAU01111	HER-2/neu protein.
6	6806	99.9	1255	20 AAU92406	Human HER-2/neu on
7	6806	99.9	1255	21 AAB21198	Human HER-2/neu pr
8	6806	99.9	1255	21 AAY84780	Amino acid sequenc
9	6806	99.9	1255	22 AAB85458	Human HER-2/neu pr
10	6806	99.9	1255	22 AAG88267	HER2/neu amino aci

11	6806	99.9	1355	23 AAE24067	Human Her-2 protei
12	6806	99.9	1355	23 AAE20479	Human Her-2/neu pr
13	6806	99.9	1355	23 AAM51143	Human Her-2/neu on
14	6806	99.9	1355	23 AAU77114	Human Her-2/neu po
15	6763	99.3	1433	14 AAR39568	Sequence of c-erbB
16	6642	97.5	1223	23 AAU98923	Human breast cance
17	6489	95.3	1300	21 AAB21208	Human HER-2/neu pr
18	5998.5	88.1	1356	21 AAB21199	Rat HER-2/neu prot
19	5998.5	88.1	1356	23 AAM51144	Rat Her-2/neu onco
20	5973.5	87.7	1356	21 AAB21206	Mouse Her-2/neu pr
21	5973.5	87.7	1356	22 AAG62860	Amino acid sequenc
22	5973.5	87.7	1356	22 AAM51151	Mouse Her-2/neu on
23	4892	71.8	919	21 AAB21203	Human HER-2/neu fu
24	4892	71.8	919	23 AAM51148	Her-2/neu extracel
25	4123.5	60.5	920	23 AAM51152	Mouse Her-2/neu ex
26	4123.5	60.5	926	23 AAM51153	Mouse HER-2/neu ex
27	3776	55.4	712	21 AAB21204	Human HER-2/neu fu
28	3776	55.4	712	23 AAM51149	Her-2/neu extracel
29	3630	53.3	782	18 AAW19764	Her2-GM-CSF immuno
30	3628	53.3	653	21 AAB21200	Extracellular HER-
31	3628	53.3	653	23 AAM51145	Human ErbB2 oncopr
32	3590	52.7	645	22 AAB60408	Human ErbB2 extrac
33	3590	52.7	645	22 AAB61593	Human ErbB2 extrac
34	3525	51.7	951	21 AAY44993	DC8ecFv-erbB2EC fu
35	3422	50.2	624	11 AAR08222	Extracellular port
36	3168	46.5	1210	21 AAB19259	Amino acid sequenc
37	3168	46.5	1210	21 AAY50616	Human EGF receptor
38	3168	46.5	1210	23 AAE23019	Human Her-1 protei
39	3168	46.5	1210	23 AAM50768	Human epidermal gr
40	3166	46.5	1210	22 AAB68420	Amino acid sequenc
41	3127	45.9	1310	23 AAB51768	Human epidermal gr
42	3110.5	45.7	654	21 AAB21205	Rat HER-2/neu prot
43	3110.5	45.7	654	23 AAM51150	Rat Her-2/neu onco
44	3084	45.3	583	23 AAE20483	Human protein for
45	3084	45.3	587	23 AAE20481	Human protein for

ALIGNMENTS

RESULT 1
AAY92620
ID AAY92620 standard; Protein; 1255 AA.
XX
AC AAY92620;
XX
DT 10-AUG-2000 (first entry)
XX
DE Human heregulin 2 (Her2).
XX
KW Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;
KW self-protein; cancer; breast cancer; prostate cancer;
KW cell-associated peptide antigen; foreign epitope.
XX
OS Homo sapiens.
XX
FH Key
FH Domain
FT /label= N-terminal
FT /note= "mature polypeptide"
FT Region
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT Region
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT Region
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT Region
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT Domain
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"

FT /label= Cysteine_rich_domain
 FT 210..224
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 250..264
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 324..483
 FT /label= Ligand_binding_domain
 FT 325..339
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 369..383
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 465..479
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 484..623
 FT /label= Cysteine_rich_domain
 FT 579..593
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 624..654
 FT /label= Transmembrane_domain
 FT 632..652
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 653..667
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 655..1010
 FT /label= Tyrosine_kinase_domain
 FT 661..675
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 695..709
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 710..730
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 1011..1235
 FT /label= C-terminal_domain
 FT
 PN W0200020027-A2.
 XX 13-APR-2000.
 XX 05-OCT-1999; 99WO-DK00525.
 XX 05-OCT-1998; 98DK-0001261.
 PR 20-OCT-1998; 98US-0105011.
 XX
 PA (MEBI-) M & E BIOTECH AS.
 XX
 PI Steinnaa L, Mouritseen S, Nielsen KG, Haaning J, Leach D, Dalum I;
 PI Gautam A, Birk P, Karlsson G;
 XX
 DR WPI: 2000-349917/30.
 DR N-PSDB; AAA09455.
 XX
 PT Inducing immune responses to weakly immunogenic, tumor associated
 PT peptide antigens for the treatment of breast and prostate cancer
 XX
 PS Claim 62; Page 193-198; 220pp; English.
 XX
 CC This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of
 CC Her2 can be used in the claimed method as an autovaccine to induce a CTL
 CC response. Subdominant CTL epitopes, antibody binding regions and
 CC cysteine residues involved in disulfide bonds are preserved in the
 CC immunogenized forms. Regions suitable for the insertion of foreign T
 CC helper epitopes were identified (see features table). The method

CC is used for inducing immune responses against weakly immunogenic
 CC cell-associated peptide antigens (PA) such as those associated with
 CC cancers (self-proteins), e.g. human prostate specific membrane antigen
 CC (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).
 CC The method comprises effecting simultaneous presentation by antigen
 CC producing cells (APCs) of the animals immune system of: (1) at least 1
 CC CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1
 CC B-cell group derived from the cell-associated PA; and (2) at least 1
 CC first T helper cell group which is foreign to the animal. Analogues of
 CC human PSM, human Her2 and human/murine FGF8b comprising a substantial
 CC part of all known and predicted CTL and B-cell epitopes of the respective
 CC PA and including at least one foreign T helper epitope are also claimed.
 CC The method is used to treat prostate, prostate/breast or breast cancer
 CC when the PA is human PSM, FGF8b and Her2, respectively.
 XX
 SQ Sequence 1255 AA;

Query Match 100.0%; Score 6812; DB 21; Length 1255;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MELAALCRWGLLLALLPPGAASSTQVCTGTDMLRLPASPTHLDMLRLHYQGVQVQGNL 60
 Db 1 MELAALCRWGLLLALLPPGAASSTQVCTGTDMLRLPASPTHLDMLRLHYQGVQVQGNL 60
 QY 61 ELTYLPTNASLSFLQDIQEVQGVLIHQNVRQVPLQRLRIVRGTFQDFEDNALAVLDNG 120
 Db 61 ELTYLPTNASLSFLQDIQEVQGVLIHQNVRQVPLQRLRIVRGTFQDFEDNALAVLDNG 120
 QY 121 DPLNNTPTVGASPGGLRELQRLSLTEILKGGVLIQORNQVPLQRLRIVRGTFQDFEDNALAVLDNG 180
 Db 121 DPLNNTPTVGASPGGLRELQRLSLTEILKGGVLIQORNQVPLQRLRIVRGTFQDFEDNALAVLDNG 180
 QY 181 LTLIDNRSRACHPCSPMCKGSRGCESEDQSLTRTVACGACRCKGPLPTDCCHEQC 240
 Db 181 LTLIDNRSRACHPCSPMCKGSRGCESEDQSLTRTVACGACRCKGPLPTDCCHEQC 240
 QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 300
 Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 300
 QY 301 YNYLSTDVGSCTLVCPHNOEVTAEADGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360
 Db 301 YNYLSTDVGSCTLVCPHNOEVTAEADGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360
 QY 361 IOEFAGCKKIFGSLAPLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLISAWPDSL 420
 Db 361 IOEFAGCKKIFGSLAPLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLISAWPDSL 420
 QY 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGISMGLRSLRELASGLALIHNNTHLCFVHTV 480
 Db 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGISMGLRSLRELASGLALIHNNTHLCFVHTV 480
 QY 481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVCNCSQFLRGQEC 540
 Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVCNCSQFLRGQEC 540
 QY 541 VVECRVLQGLPREYVNAHCLPCHPECQPNQSVTCFGEADQCACAHYKDPFCVCAR 600
 Db 541 VVECRVLQGLPREYVNAHCLPCHPECQPNQSVTCFGEADQCACAHYKDPFCVCAR 600
 QY 601 PSGVKPDLSPMPLWPKFPDEEGACQPCPINCTHSCVDLDDKGCPEAQASPLTSIVSAVVG 660
 Db 601 PSGVKPDLSPMPLWPKFPDEEGACQPCPINCTHSCVDLDDKGCPEAQASPLTSIVSAVVG 660
 QY 661 ILLVVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720
 Db 661 ILLVVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720
 QY 721 RKVKVLGSGAFCTGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVGP 780
 Db 721 RKVKVLGSGAFCTGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVGP 780

Qy	* 781	YVSELLGICLTSTVQIUTQLMYPYCCLLDHVRENKRGISQDLLNWCQIAKMSYLEDVR	840
Db	781	YVSELLGICLTSTVQIUTQLMYPYCCLLDHVRENKRGISQDLLNWCQIAKMSYLEDVR	840
Qy	841	LVHRDLAARNVLKSPNHVKITDIFGLARLLDIDETEHADGGKVPINWMALESILRRFT	900
Db	841	LVHRDLAARNVLKSPNHVKITDIFGLARLLDIDETEHADGGKVPINWMALESILRRFT	900
Qy	901	HQSDVWSYGVTWVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMVVKCWM	960
Db	901	HQSDVWSYGVTWVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMVVKCWM	960
Qy	961	IDSECRPRFRELVSERFARMARDPORFVVIQNEIDLGPASPLDSTFYRSLLEDDDMGDVLDA	1020
Db	961	IDSECRPRFRELVSERFARMARDPORFVVIQNEIDLGPASPLDSTFYRSLLEDDDMGDVLDA	1020
Qy	1021	EYVLVPQQGFCDPAPGAGVMVHRRHSSTSRSGGDLTLGLEPSEEEAPRSPAPSEG	1080
Db	1021	EYVLVPQQGFCDPAPGAGVMVHRRHSSTSRSGGDLTLGLEPSEEEAPRSPAPSEG	1080
Qy	1081	AGSDVFDGDLGMGAAGKGLQSLPTHDPSPLORYSEDPVTVLPSETDGVVAPLTCSPQPEYV	1140
Db	1081	AGSDVFDGDLGMGAAGKGLQSLPTHDPSPLORYSEDPVTVLPSETDGVVAPLTCSPQPEYV	1140
Qy	1141	NQSDVREPQPSREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAGGAVENPEYLTQ	1200
Db	1141	NQSDVREPQPSREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAGGAVENPEYLTQ	1200
Qy	1201	GGAAPOHPHPPAPSPAFDNLVYWDQDDPPERGAPPESTFKGTPTAENPYLGLDVVP	1255
Db	1201	GGAAPOHPHPPAPSPAFDNLVYWDQDDPPERGAPPESTFKGTPTAENPYLGLDVVP	1255

RESULT 2	
AAE12130	
ID	AAE12130 standard; Protein; 1255 AA.
XX	
AC	AAE12130;
XX	
DT	18-DEC-2001 (first entry)
XX	
DE	Human tyrosine kinase-type receptor, HER-2.
XX	
KW	Therapeutic compound; major histocompatibility complex; vaccine;
KW	antigenic peptide; MHC; immunoregulatory; immune response; HER-2;
KW	adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;
KW	antigen presenting cell; human; tyrosine kinase-type receptor.

XX	Key	Location/Qualifiers
XX	FH	774..782
FT	Region	/note= "Antigenic epitope"
FT		
XX	XX	WO200168677-A2.
XX	PN	
XX	PN	
XX	PD	
XX	PD	20-SEP-2001.
XX	XX	
PF	PF	16-MAR-2001; 2001WO-US40328.
XX	XX	
XX	XX	16-MAR-2000; 2000US-0527487.
XX	XX	
XX	XX	(GENZ) GENZYME CORP.
XX	PI	
XX	PI	Nicolette CA;
XX	XX	
XX	WFI	2001-616284/71.
DR	N-PSDS	AAD19731.
DR	N-PSDS	

PT Novel synthetic therapeutic compound for inducing immune response and
PT for use in adoptive immunotherapy, has enhanced binding to major
PT histocompatibility molecules and enhanced immunoregulatory properties
PT .

Claim 4; Page 63-67; 69pp; English.

The invention relates to synthetic therapeutic compounds (antigenic peptides) with enhanced binding to major histocompatibility complex (MHC) molecules and enhanced immunoregulatory properties relative to their natural counterparts. Compounds of the invention are useful for inducing an immune response in a subject and for use in adoptive immunotherapy. They are useful as components of anti-cancer vaccines and to expand immune effector cells that are specific for cancers characterised by expression of the breast cancer antigen, HER-2. Polynucleotides that encode peptides of the invention are useful as hybridisation probes and as primers for the detection of genes of genes transcripts that are expressed in antigen presenting cells (APCs), to confirm transduction of polynucleotides into host cells. The present sequence is human tyrosine kinase-type receptor, HER-2. Compounds of the invention are designed based on the HER-2 antigenic peptide (774-782).

Q	SQ	Sequence	1255 AA;	
		Query Match	100.0%; Score 6812; DB 22; Length 1255;	
		Best Local Similarity	100.0%; Pred. No. 0;	
		Matches 1255; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
Qy	1	MELAAALCRWGILLALLAPP	GGAASSTQVCTGTDWKLRLPASPETHLDMLRHLYGQCQVVG	60
Db	1	MELAAALCRWGILLALLAPP	GGAASSTQVCTGTDWKLRLPASPETHLDMLRHLYGQCQVVG	60
Qy	61	ELTYLPTNASLSFLQDIOEQVGYVLI	IAHNQVRQVPLQRLRIVRGTQFLFEDNYALAVLDNG	120
Db	61	ELTYLPTNASLSFLQDIOEQVGYVLI	IAHNQVRQVPLQRLRIVRGTQFLFEDNYALAVLDNG	120
Qy	121	DPLNNTTPVTGASPGGLRELQ	LSLTIELKGGVLIQRNPOLCYQDITLWKDI	180
Db	121	DPLNNTTPVTGASPGGLRELQ	LSLTIELKGGVLIQRNPOLCYQDITLWKDI	180
Qy	181	LTLLDITNRSRACHPCSPMCKG	SRGWSSSDCOSLTRTVCAGGCARCKGPLPTDCCHBQC	240
Db	181	LTLLDITNRSRACHPCSPMCKG	SRGWSSSDCOSLTRTVCAGGCARCKGPLPTDCCHBQC	240
Qy	241	AAGCTGPKHSDCLACLHFNHSG	ICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP	300
Db	241	AAGCTGPKHSDCLACLHFNHSG	ICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP	300
Qy	301	YNYLSTDVVGCTLVCP	PLHNQEVTAEDGTQCEKSKPCARVCYGLGMEHLREVR	360
Db	301	YNYLSTDVVGCTLVCP	PLHNQEVTAEDGTQCEKSKPCARVCYGLGMEHLREVR	360
Qy	361	IQEFAGCKKIFGSLAF	PESFDGDPASNTAPLQPEQLQVFTTEIITGYLYISAWPDSLP	420
Db	361	IQEFAGCKKIFGSLAF	PESFDGDPASNTAPLQPEQLQVFTTEIITGYLYISAWPDSLP	420
Qy	421	DLVSFQNLQVIRGRILHNGAYS	ITLQGLGTSWLGLSRLRELGSGALIHNNTHLCFVHTV	480
Db	421	DLVSFQNLQVIRGRILHNGAYS	ITLQGLGTSWLGLSRLRELGSGALIHNNTHLCFVHTV	480
Qy	481	PWQLFPNPHQALLHTANR	PEDECVGSGLACHQLCARGHCWGPPTOCVNCQSQFLRGQEC	540
Db	481	PWQLFPNPHQALLHTANR	PEDECVGSGLACHQLCARGHCWGPPTOCVNCQSQFLRGQEC	540
Qy	541	VEECRVLQGLPREYV	NARHCLPCHPECPQNGSVTCFGEADQCACAHYKDPDFCVARC	600
Db	541	VEECRVLQGLPREYV	NARHCLPCHPECPQNGSVTCFGEADQCACAHYKDPDFCVARC	600
Qy	601	PSGVKPDLSYMPITWK	SPDEBGAQCPDINTHSCVDLDDKGCPAEQASPLT	660
Db	601	PSGVKPDLSYMPITWK	SPDEBGAQCPDINTHSCVDLDDKGCPAEQASPLT	660
Qy	661	ILLVVLGVVFGILIKR	QOQIKRYTNRLLIQTETLVEPLTPSGAMPNQAMRILKETEL	720
Db	661	ILLVVLGVVFGILIKR	QOQIKRYTNRLLIQTETLVEPLTPSGAMPNQAMRILKETEL	720

QY 721 RKVKVLGSGAGFYVYGIWIPDGENVKIPVAIKVLRENTSPKANKIIDEAYMAGVGP 780
 DB 721 RKVKVLGSGAGFYVYGIWIPDGENVKIPVAIKVLRENTSPKANKIIDEAYMAGVGP 780
 QY 781 YVSRLLGICLTSTVQLVTLQMPYGCCLLDHVRNRRGLSGDGLLNCWQIAKMSYLEVDV 840
 DB 781 YVSRLLGICLTSTVQLVTLQMPYGCCLLDHVRNRRGLSGDGLLNCWQIAKMSYLEVDV 840
 QY 841 LVHRDLAARNVLKSPNHNKIDPGLARLLDIDETEVHADGGKVPKMWALESIILRRFT 900
 DB 841 LVHRDLAARNVLKSPNHNKIDPGLARLLDIDETEVHADGGKVPKMWALESIILRRFT 900
 QY 901 HQSDVMSYGYVWELMTFGAKPYDGPAREIPDLLEKGERLPPOPICTIDVYIMVWKWM 960
 DB 901 HQSDVMSYGYVWELMTFGAKPYDGPAREIPDLLEKGERLPPOPICTIDVYIMVWKWM 960
 QY 961 IDSECRPRFRELVSERFMRARDQRFVVIQNEIDLGSPASPLDSTFYRSLLEDDBDGLVDA 1020
 DB 961 IDSECRPRFRELVSERFMRARDQRFVVIQNEIDLGSPASPLDSTFYRSLLEDDBDGLVDA 1020
 QY 1021 EYLVPOQGFCDPAPAGMGVHHRSSSTSGGDLTLGLEPSEEAAPRPLAPSEG 1080
 DB 1021 EYLVPOQGFCDPAPAGMGVHHRSSSTSGGDLTLGLEPSEEAAPRPLAPSEG 1080
 QY 1081 AGSDVFDGDLGMAAGLQSLPDLTHPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
 DB 1081 AGSDVFDGDLGMAAGLQSLPDLTHPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
 QY 1141 NQPDVPRQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAGGAVENPEYLTPO 1200
 DB 1141 NQPDVPRQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAGGAVENPEYLTPO 1200
 QY 1201 GGAAPQHPPPAPSPADNLYYWDQPPPERGAPSTFKGTPTAENPEYLGLDVPU 1255
 DB 1201 GGAAPQHPPPAPSPADNLYYWDQPPPERGAPSTFKGTPTAENPEYLGLDVPU 1255

RESULT 3

AAB60167

ID AAB60167 standard; Protein; 1255 AA.

XX AC AAB60167;

XX DT 03-APR-2001 (first entry)

XX DE HER2 transgene plasmid construct encoded protein.

XX KW Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;

XX OS antibody.

XX OS Homo sapiens.

XX OS Synthetic.

XX PN WO200100244-A2.

XX PD 04-JAN-2001.

XX PF 23-JUN-2000; 2000WO-US17229.

XX PR 25-JUN-1999; 99US-0141316.

XX PR 16-MAR-2000; 2000US-0189844.

XX PA (GETH) GENENTECH INC.

XX PI Erickson S, Schwall R;

XX DR WPI; 2001-061962/07.

XX DR N-PSDB; AAF24297.

XX PT Treating tumors, particularly breast cancers, which overexpress an ErbB

XX PT receptor and does not respond to an anti-ErbB antibody, comprises

XX PT conjugating the antibody to a maytansinoid -

XX PT

PS Example 3; Fig 4; 92pp; English.
 CC The present invention provides a method of treating cancer by
 CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In
 CC particular, the antibody is directed against ErbB2 (also known as HER2
 CC and p185neu). The method is particularly useful in the treatment of
 CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
 CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.
 XX Sequence 1255 AA;

	Query Match	Best Local Similarity	Score	DB	Length	1255;
	Matches	Conservative	0;	Mismatches	0;	Indels
	0;	0;	0;	0;	0;	Gaps
QY	1	MELAAALCRWGLLALLPPGAAS	100.0%;	100.0%;	6812;	DB 22;
DB	1	MELAAALCRWGLLALLPPGAAS	100.0%;	100.0%;	6812;	DB 22;
QY	61	ELTYLPTNASLSFLQDIQEVQ	100.0%;	100.0%;	6812;	DB 22;
DB	61	ELTYLPTNASLSFLQDIQEVQ	100.0%;	100.0%;	6812;	DB 22;
QY	121	DFLNNTPVTGASPGSLRELQ	100.0%;	100.0%;	6812;	DB 22;
DB	121	DFLNNTPVTGASPGSLRELQ	100.0%;	100.0%;	6812;	DB 22;
QY	181	LFLIDTNRSRACHSPCKSG	100.0%;	100.0%;	6812;	DB 22;
DB	181	LFLIDTNRSRACHSPCKSG	100.0%;	100.0%;	6812;	DB 22;
QY	241	AAGCTGPKHSDCLACLFH	100.0%;	100.0%;	6812;	DB 22;
DB	241	AAGCTGPKHSDCLACLFH	100.0%;	100.0%;	6812;	DB 22;
QY	301	YNLSTDVGSCTLVCPHNE	100.0%;	100.0%;	6812;	DB 22;
DB	301	YNLSTDVGSCTLVCPHNE	100.0%;	100.0%;	6812;	DB 22;
QY	361	IQEFAGCKIKFSLAPLPS	100.0%;	100.0%;	6812;	DB 22;
DB	361	IQEFAGCKIKFSLAPLPS	100.0%;	100.0%;	6812;	DB 22;
QY	421	DLVSFONLOVIRGRILHN	100.0%;	100.0%;	6812;	DB 22;
DB	421	DLVSFONLOVIRGRILHN	100.0%;	100.0%;	6812;	DB 22;
QY	481	PWDQLFRNPHQALLHTAN	100.0%;	100.0%;	6812;	DB 22;
DB	481	PWDQLFRNPHQALLHTAN	100.0%;	100.0%;	6812;	DB 22;
QY	541	VEECRVLOGLPREYVNAH	100.0%;	100.0%;	6812;	DB 22;
DB	541	VEECRVLOGLPREYVNAH	100.0%;	100.0%;	6812;	DB 22;
QY	601	PSGVKPDLSYMPITWKF	100.0%;	100.0%;	6812;	DB 22;
DB	601	PSGVKPDLSYMPITWKF	100.0%;	100.0%;	6812;	DB 22;
QY	661	ILLVVLGVVFGILIKRR	100.0%;	100.0%;	6812;	DB 22;
DB	661	ILLVVLGVVFGILIKRR	100.0%;	100.0%;	6812;	DB 22;
QY	721	RKVKVLGSGAGFYVYGIW	100.0%;	100.0%;	6812;	DB 22;
DB	721	RKVKVLGSGAGFYVYGIW	100.0%;	100.0%;	6812;	DB 22;
QY	781	YVSRLLGICLTSTVQLV	100.0%;	100.0%;	6812;	DB 22;
DB	781	YVSRLLGICLTSTVQLV	100.0%;	100.0%;	6812;	DB 22;
QY	841	LVHRDLAARNVLKSPNH	100.0%;	100.0%;	6812;	DB 22;
DB	841	LVHRDLAARNVLKSPNH	100.0%;	100.0%;	6812;	DB 22;

Db * 841 LVHRDLAARNVLYKSPNHVKITDFGLARLLIDIDETEHADGGKVPKMALESILRRPT 900

Qy 901 HQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICITIDVYIMVKCWM 960

Db 901 HQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICITIDVYIMVKCWM 960

Qy 961 IDSECRPRFRELVSERFARMARDQRFVVIQNEIDLGRASPLDSTFYRSLLEDMDGDLVDA 1020

Db 961 IDSECRPRFRELVSERFARMARDQRFVVIQNEIDLGRASPLDSTFYRSLLEDMDGDLVDA 1020

Qy 1021 EYVLVPOQGFCCPDPAFCAGGVHHRSSSTSSGGGDLTLGLEPSEEEAPRPLAPSEG 1080

Db 1021 EYVLVPOQGFCCPDPAFCAGGVHHRSSSTSSGGGDLTLGLEPSEEEAPRPLAPSEG 1080

Qy 1081 AGSDVFDGDLGMAAKGLQSLPHTDPSPLQKYSIEDPTVLPSETDGYVAPLTCSPQPEYV 1140

Db 1081 AGSDVFDGDLGMAAKGLQSLPHTDPSPLQKYSIEDPTVLPSETDGYVAPLTCSPQPEYV 1140

Qy 1141 NOPDVRPQPSREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAGGAVENPEYLTPO 1200

Db 1141 NOPDVRPQPSREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAGGAVENPEYLTPO 1200

Qy 1201 GGAAPQHPPPAFAFDNLVYWDQDPPERGAPPSTPKGTPTAENPEYLGLDVVP 1255

Db 1201 GGAAPQHPPPAFAFDNLVYWDQDPPERGAPPSTPKGTPTAENPEYLGLDVVP 1255

RESULT 4

AAU74545

ID AAU74545 standard; Protein; 1255 AA.

XX AAU74545;

DT 23-APR-2002 (first entry)

XX Human HER2 (ErbB2) polypeptide.

XX Human; HER2; ErbB; epidermal growth factor receptor; receptor;

KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;

KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;

KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;

KW glial disorder; astrocytal disorder; hypothalamic disorder;

KW glandular disorder; macrophagal disorder; epithelial disorder;

KW stromal disorder; blastocoealic disorder; inflammatory disorder;

KW angiogenic disorder; immunological disorder.

XX Homo sapiens.

XX OS

XX US2002001587-A1.

XX 03-JAN-2002.

XX 16-MAR-2001; 2001US-0811123.

XX 16-MAR-2000; 2000US-189844P.

PR 05-OCT-2000; 2000US-238327P.

XX (ERIC/) ERICKSON S.

PA (SCHW/) SCHWALL R.

PA (SLIW/) SLIWOWSKI M.

XX Erickson S, Schwall R, Sliwowski M;

PI WPI; 2002-163686/21.

XX N-PSDB; ABK14058.

DR DR

XX Treating tumour characterised by overexpression of epidermal growth

PT factor receptor, ErbB or cancer in mammal, comprises administering

PT anti-ErbB antibody-maytansinoid conjugate to the mammal

XX Example 3; Fig 7; 93pp; English.

PS The invention relates to treating a tumour in a mammal, where the tumour

CC

CC is characterised by the overexpression of an epidermal growth factor

CC receptor (ErbB) and does not respond or responds poorly, to treatment

CC with an anti-ErbB antibody, comprising administering to the mammal an

CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for

CC treating cancer or tumours of the breast, ovary, stomach, endometrium,

CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,

CC prostate and bladder, preferably breast cancer. The breast cancer is a

CC metastatic breast cancer or an aggressive form of metastatic breast

CC cancer which overexpresses ErbB2. The method is also useful for treating

CC neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal,

CC epithelial, stromal, blastocoealic, inflammatory, angiogenic and

CC immunological disorders. This sequence represents the human HER2 (ErbB2)

CC polypeptide of the invention.

XX

XX Sequence 1255 AA;

Query Match 100.0%; Score 6812; DB 23; Length 1255;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLALLPPGAASQTQCTGDMKRLRSPASPETHLDMLRHLYQGCVVQGNL 60

Db 1 MELAALCRWGLLALLPPGAASQTQCTGDMKRLRSPASPETHLDMLRHLYQGCVVQGNL 60

Qy 61 ELYLPTNASLSFLQDIEVQGVYLIHNNQVQVPLQRLRIVRGTOIFEDNYALAVLDNG 120

Db 61 ELYLPTNASLSFLQDIEVQGVYLIHNNQVQVPLQRLRIVRGTOIFEDNYALAVLDNG 120

Qy 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFKHKNOLA 180

Db 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFKHKNOLA 180

Qy 181 LTLIDTNRSRACHPCSPMCKGSRGWESSEDCQSLTRTVCCAGCARCKGPLPTDCCHEQC 240

Db 181 LTLIDTNRSRACHPCSPMCKGSRGWESSEDCQSLTRTVCCAGCARCKGPLPTDCCHEQC 240

Qy 241 AGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPBGRVTFGASCVTACP 300

Db 241 AGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPBGRVTFGASCVTACP 300

Qy 301 YNYLSDVGSCTLVCPHNOEVTAEQDQRCCKPCARVCYGLGMEHLREVRVTSAN 360

Db 301 YNYLSDVGSCTLVCPHNOEVTAEQDQRCCKPCARVCYGLGMEHLREVRVTSAN 360

Qy 361 IOEFAGCKKIFGSLAPLPSFDGDPASNTAPLQPEQLQVFETILEETGYLISAWPDSLP 420

Db 361 IOEFAGCKKIFGSLAPLPSFDGDPASNTAPLQPEQLQVFETILEETGYLISAWPDSLP 420

Qy 421 DLSVFONLQVIRGRILHNGAYSILTQGLGISWGLRSLRSLRELGSGLALIHNTHLCFVHTV 480

Db 421 DLSVFONLQVIRGRILHNGAYSILTQGLGISWGLRSLRSLRELGSGLALIHNTHLCFVHTV 480

Qy 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHOLCARGHCWGPGTQCVCNCSQFLRGQEC 540

Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHOLCARGHCWGPGTQCVCNCSQFLRGQEC 540

Qy 541 VBECRVLQGLPZYVNNARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600

Db 541 VBECRVLQGLPZYVNNARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600

Qy 601 PSGVKPDLISYMPIWKFPDDEGACQPCPNCTHSCVDLDDKGCPCAPQASPLTSIVSAVVG 660

Db 601 PSGVKPDLISYMPIWKFPDDEGACQPCPNCTHSCVDLDDKGCPCAPQASPLTSIVSAVVG 660

Qy 661 ILLVWVVGVLVIRGRILKRRQKIRKVTMRLLQFTELVEPLTPSGAMPNQAQRILKETEL 720

Db 661 ILLVWVVGVLVIRGRILKRRQKIRKVTMRLLQFTELVEPLTPSGAMPNQAQRILKETEL 720

Qy 721 RKVKVLGSAFGTGYVKGIVIPGENVKIPVALKVLRENTSPKANKEILDEAYMAGVGSF 780

Db 721 RKVKVLGSAFGTGYVKGIVIPGENVKIPVALKVLRENTSPKANKEILDEAYMAGVGSF 780

Qy 781 YVSRLLGICLTSTVQLVTQJLMPYGCILLDHVRENRLRGLSQDILLNWCMTAKGMSYLEDYR 840

Db 781 YVSRLLGICLTSTVQLVTPYGLDHRNRRGLSGDGLLNCWQIAKMSYLEDDR 840
Qy 841 LVHRDLAARNVLKSPNHVITDFGLARLLIDIDETEHADGGKVPKMALESILRRFT 900
Db 841 LVHRDLAARNVLKSPNHVITDFGLARLLIDIDETEHADGGKVPKMALESILRRFT 900
Qy 901 HQSDWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPICTIDVYIMVWKCM 960
Db 901 HQSDWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPICTIDVYIMVWKCM 960
Qy 961 IDSECRPRFRELVSERFARMARDPQFVVIQNEIDLGPAASPLDSTFYRSLLEDGMDLDA 1020
Db 961 IDSECRPRFRELVSERFARMARDPQFVVIQNEIDLGPAASPLDSTFYRSLLEDGMDLDA 1020
Qy 1021 EYLVPOQGFCDPAPAGAGVHRRSSSTRSGGDLTLGLEPSEERAPRPLAPSEG 1080
Db 1021 EYLVPOQGFCDPAPAGAGVHRRSSSTRSGGDLTLGLEPSEERAPRPLAPSEG 1080
Qy 1081 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
Qy 1141 NQPDVPPQSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAGGAVENPEYLTPO 1200
Db 1141 NQPDVPPQSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAGGAVENPEYLTPO 1200
Qy 1201 GGAAPQHPHPPAPAFDNLVYWDQDPPRGAPPSTFKGTPTAENPEYLGIDVPV 1255
Db 1201 GGAAPQHPHPPAPAFDNLVYWDQDPPRGAPPSTFKGTPTAENPEYLGIDVPV 1255

RESULT 5

AAW01111 standard; Protein; 1255 AA.
XX AC AAW01111;
DT 01-JAN-1997 (first entry)
XX DE HER-2/neu protein.
KW HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
KW breast cancer; ovary cancer; colon cancer; lung cancer;
KW prostate cancer; immunisation; tumour; vaccine; vector.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT Domain 676..1255
FT /label= Intracellular domain
FT /note= "Claimed domain, useful for immunisation"
XX PN W09630514-A1.
XX PD 03-OCT-1996.
XX PF 28-MAR-1996; 96WO-US01689.
XX PR 31-MAR-1995; 95US-0414417.
XX PA (UNIW) UNIV WASHINGTON.
XX PI Cheever MA, Disis ML;
XX DR WPI; 1996-455361/45.
XX DR N-PSDB; AAT40739.
XX PT DNA encoding-HER-2-neu poly-peptide(s) - used for prevention or
XX PT treatment of malignancies with which the HER-2/neu oncogene is
XX PT associated
XX PS Claim 2; Page 56-61; 71pp; English.

XX Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is
CC the product of the HER-2/neu oncogene (see also AAT40739). The
CC protein is over-expressed in various cancers, including breast,
CC ovarian, colon, lung and prostate. The intracellular domain of the
CC protein can be used to immunise an animal against a malignancy with
CC which the oncogene is associated. The polypeptide can be produced
CC in transformed host cells for use in immunisation. Alternatively,
CC animal cells are transfected in vivo or ex vivo with a viral vector
CC that directs expression of the polypeptide.
XX SQ Sequence 1255 AA;
Query Match 99.9%; Score 6806; DB 17; Length 1255;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1253; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MELAALCRWGLLLALLPPGAASSTQVCTGDMKRLPASPETHLDMRLHLYQGCVVQGNL 60
Db 1 MELAALCRWGLLLALLPPGAASSTQVCTGDMKRLPASPETHLDMRLHLYQGCVVQGNL 60
Qy 61 ELYLPTNASLSFLQDIQEVQVLIHNVQVPLQRLRIVRGTOIFEDNYALAVLNDG 120
Db 61 ELYLPTNASLSFLQDIQEVQVLIHNVQVPLQRLRIVRGTOIFEDNYALAVLNDG 120
Qy 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIHFKNNOLA 180
Db 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIHFKNNOLA 180
Qy 181 LTLIDNTRSRACHPCSPMCKGSRGWESSEDCQSLTRTVAGCARGKGLPTDCCHEQC 240
Db 181 LTLIDNTRSRACHPCSPMCKGSRGWESSEDCQSLTRTVAGCARGKGLPTDCCHEQC 240
Qy 241 AGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMNPGRVTFGASCVTACP 300
Db 241 AGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMNPGRVTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPLNHNEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNYLSTDVGSCTLVCPLNHNEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360
Qy 361 IQEFAGCKKIFGSLAFPLPSFDGDPASNTAPLQEQVLFETLEETGLYISAWPDSL 420
Db 361 IQEFAGCKKIFGSLAFPLPSFDGDPASNTAPLQEQVLFETLEETGLYISAWPDSL 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSLLTQGLGSLGLSLRELGLALIHNTLHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSLLTQGLGSLGLSLRELGLALIHNTLHLCFVHTV 480
Qy 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCGPQTQCVNCSQFLRGQEC 540
Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCGPQTQCVNCSQFLRGQEC 540
Qy 541 VEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFVCAR 600
Db 541 VEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFVCAR 600
Qy 601 PSQVKPDLSPYMPIWKFPEDEGACQPCINCHTSCVDLDDKGCFAEORASPLTSIVSAVVG 660
Db 601 PSQVKPDLSPYMPIWKFPEDEGACQPCINCHTSCVDLDDKGCFAEORASPLTSIVSAVVG 660
Qy 661 ILLVVVLGVVFGILIKRQOKIRKYMRLLOSTELVEPLTPSGAMPNOAOMRILKETEL 720
Db 661 ILLVVVLGVVFGILIKRQOKIRKYMRLLOSTELVEPLTPSGAMPNOAOMRILKETEL 720
Qy 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVSP 780
Db 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVSP 780
Qy 781 YVSRLLGICLTSTVQLVTPYGLDHRNRRGLSGDGLLNCWQIAKMSYLEDDR 840
Db 781 YVSRLLGICLTSTVQLVTPYGLDHRNRRGLSGDGLLNCWQIAKMSYLEDDR 840

Db 772 LLGLCGSSQLQVTVQLPLGSLLDHVKQHRETLGPQLLLNMGVQIAKGMYYLBEHSMVHR 831
QY 839 DLARNVVKSPNHNKVTDFGLARLLDIDETEHADGKGKVPKWKMALESILRRRFTHQSD 898
Db 832 DLALRNWMLKSPSQVQVADFGVADLLPPDDKQLLHSEAKTFIKWMALESIFHCYKTHQSD 891
QY 899 VWSYGVTVWELMTFGAKPYDGIAPRELIPDLLEKGERLPQPPICITIDVYMWKMWKMWIDSE 958
Db 892 VWSYGVTVWELMTFGAEPYAGRLAEIPDLLEKGERLAQPOICITIDVYMWKMWKMWIDEN 951
QY 959 CRPRFELVSEFRMARDPQRFVVIQNEGLPASPDLSTFYRSLLEDDDDMGDLVDAEYVL 1018
Db 952 IRTTFELANEFTRMARDPPYLVIKRAS-GPQTP--PAAPSPVITTKEL-----QAEEL 1003
QY 1019 VPOQGFCCPDPAFGAGGMVHRRSSSTRSGGDLTLGLPSEB----- 1062
Db 1004 EPEL-----DLDLLEAEELGATSLGALSLSPTGT 1034
QY 1063 -----EAPRSLAPSEG-----AGSDVPDGLGMAAGKGLSLPHTDPSLPQRYSEDP 1110
Db 1035 LTRPRGSQSLSPSSGYMPMNQSSLGACLDLSAVLGGREQFSRPISLH-PIPRGR----- 1088
QY 1111 TVPLPSETGVY---APL-----TC-----SPOPE-----YVNOPDVRPQPSPREGP 1150
Db 1089 --PASESSEGHVTCSEAELOEKVSVCRSRSSRSPRGRSATHSRHSLTTPVTPLSP 1146
QY 1151 -----LPAARPAGATLRAKTLSP-KGNQVY-----KDVFAFGGAVENPEYLT 1192
Db 1147 GLSEEDGNGVYMPDTHLRGASSREGTLSSVGLSSVLGTDEED-----EYFYMN 1198
QY 1193 PQGGAAPQPHPP 1204
Db 1199 RKRGSP-PRPP 1209
RESULT 11
TVFVLV
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian leukosis virus
N;Contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase
C;Species: avian leukosis virus, ALV
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 11-Jun-1999
C;Accession: B00643; A00643
R;Nilsson, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1995
A;Title: C-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro-
A;Reference number: A00643; MUID:85228222; PMID:2988784
A;Accession: B00643
A;Molecule type: mRNA
A;Residues: 1-698 <NLT>
A;Cross-references: GB:M10066; GB:M13881; NID:G211749; PIDN:AAA48763.1; PID:G211750
A;Note: in Genbank entry CHKERBBF, release 109.0, the source is designated as Gallus gal
C;Comment: This protein is synthesized as a gag-env-erbB protein.
C;Genetics:
A;Gene: gag-env-erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific p
F;1-6/Product: gag protein (fragment) #status predicted <GAG>
F;7-59/Product: env protein (fragment) #status predicted <ENV>
F;60-698/Product: protein-tyrosine kinase erbB #status predicted <ERB>
F;194-459/Domain: protein kinase homology <KIN>
F;202-210/Region: protein kinase ATP-binding motif
F;229/Active site: Lys #status predicted

Query Match 26.18; Score 1766.5; DB 1; Length 698;
Best Local Similarity 52.24; Pred. No. 2.3e-65;
Matches 374; Conservative 80; Mismatches 137; Indels 125; Gaps 18;

QY 572 GPRADQCVACAHYKDPFPCVAPCGSVKPDLSYMPIWKFPDEGACQPCPINCTHSCVDL 631
Db 60 GP--DHCWKAHFIDGPHCVKACPAGVLGENDTL-VWKYADANAVCOLCHPNCRTGCKGP 116
QY 632 DDKGCPRAEQASPLTSTVSAVV-GILLVVVLGVVFGILIKRQOKTKRYTMRLLQSTEL 690

Db 117 GLEGCP---NGSKTFSIAAGVVGGLLCLVVVGLGIGLYLRRR-HIVRKETLRLLOEREL 172
QY 691 VEPLTPSGAMPNQAMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVL 750
Db 173 VEPLTPSGEAPNQAHLRIILKETEFKKVKVLGSGAFGTVYKGLWIPGEKVKIPVAIKEL 232
QY 751 ENTSPKANKEIIDEAAYMAGVCSPPVSRLLGLCLTSTVQLVTLQMLPYGCLLDHVRNRR 810
Db 233 EATSPKANKEIIDEAAYWASVDNPHVCRLLGLCLTSTVQLITQLMPYGCGLLDYIREHKN 292
QY 811 LGSQDLNWCMIQAKMSYLEVRLVHRDLAARNVLKSPNHNKVTDFGLARLLDIDETE 870
Db 293 IGSQVLLNWCMIQAKGMVLEERLVRDLAARNVLKTPQHVKITDFGLAKLLGADEKE 352
QY 871 YHADGKVP:KMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIAPAREIPDLLE 930
Db 353 YHAEKGKVP:KMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIAPASEISVLE 412
QY 931 KGERLPQPPICITIDVYMWKMWKMWIDSECRPPRELVSFSESMARDPQRFVVIQ-NEDLG 989
Db 413 KGERLPQPPICITIDVYMWKMWKMWIDADSRPKFERLIAEFSKWARDPPRYLVIGQDERMH 472
QY 990 PASPLDSTVRSILLEDDMGDLVDAEYLVPOQGFCDPAFGAGGMVHRRSSSTRSG 1049
Db 473 LESPDSKFTYRIMEEDMEDIVDAEYLVPHQGF-----NSPST----- 513
QY 1050 GGDLTGLPSEBEEAPRSP-----APSEAGSDVPDGLGMAAGKGLSLPHTDPSPLQ 1104
Db 514 -----SRTELLSSLSATSNNSATCID-----RNGQGHFVRESFVQ 550
QY 1105 RYSEDDTVPILPSET--DGIVAPLTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAGATLE 1162
Db 551 RYSSDDPTGNFLBESIDDDGFL-----PAPEYVQ--LMPKPS----- 585
QY 1163 RAKTSLSPGKGVVQKVF-----AFGAVENPEYLTPOGGAAPQPHPPAF 1207
Db 586 ----TAMVQNIYNNISLTAISKLPMSRYQNSHSTAVDNPEYL-----NTNOSPLA 633
QY 1208 SPAFNLYYWDQ-----DPPE-----RGAPPSTFKGTPTAENPEYLGHDVP 1248
Db 634 KTVFESSPYWIOGNHQLNDNPVQDDFLPNETKENGLLKVPAAENPEYLVRAAP 689
RESULT 12
TVYUHV
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain H)
C;Species: avian erythroblastosis virus
C;Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 11-Jun-1999
C;Accession: A00644; A38022
R;Yamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.
Cell 35, 71-78, 1983
A;Title: The erbB gene of avian erythroblastosis virus is a member of the src gene fami
A;Reference number: A00644; MUID:84026539; PMID:6313229
A;Accession: A00644
A;Molecule type: DNA
A;Residues: 1-604 <YAM>
A;Cross-references: GB:K01216; NID:G209676; PIDN:AAA42400.1; PID:G209678
R;Debuire, B.; Henry, C.; Benaissa, M.; Biserte, G.; Claverie, J.M.; Saule, S.; Martin,
Science 224, 1456-1459, 1984
A;Title: Sequencing the erbA gene of avian erythroblastosis virus reveals a new type of
A;Reference number: A38022; MUID:84223957; PMID:6328658
A;Accession: A38022
A;Molecule type: DNA
A;Residues: 1-28,'W',30-139,'F',141-145,'V',147-152 <DEB>
A;Cross-references: GB:K02006
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific
F;130-395/Domain: protein kinase homology <KIN>
F;138-146/Region: protein kinase ATP-binding motif
F;165/Active site: Lys #status predicted


```

Db      715 QYTAIGPY-----CAASPPRSKITANLD-----VNMIFITGAVLVPTIC 755
QY      663 VVFGI-LIKRQOKIRKYT--MRLLQETELVBLTPSGAMPNQAQMRILKETELRKVKV 719
Db      756 ILCVWTYICROKQKAKETVMTMALSGREDSPLPSNIGANLKLIRVKDAELRGGV 815
QY      720 LGSAGRTYVKGWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVGSPPVSR 779
Db      816 LGMAGFRYVKGWVPEGENVKIPVAIKELLKTSAGSESEFEUREAYIMASEEHVLLKL 875
QY      780 LGICLSTVOLVTQMLPYGCLLDHVRNGLSGQDLLANCMQIAKGMVLELDVRLVHRD 839
Db      876 LAVCMSQMWLITQLMPLGCLLDYVRNRDKIGSKALLNWSQIAKGMVLEEKRLVHRD 935
QY      840 LAARNVLVK--SPNHVKITDFGLARLDIDETEHADGGKVPKIMWALESILRRFFTHQ 896
Db      936 LAARNVLVRLAGEDH---DFGLAKLLSSDSNEYKAAGGKMPKWLALCEIRNRVTSK 991
QY      897 SDVWSGVTVWELMTGAKPYDGPAREIPDLLEKGERLPPOPICTIDVYIMVVKWMD 956
Db      992 SDVWARGVTIWELLTFQRPHEINIPAKDIPDLIEVGLKLEQPBICSLDIYCTLLSCWILD 1051
QY      957 SECRPRFELVSEFSRMDPQRFVVIQNEIDLQ--PASPLDSTFYRSLLEDD---DMGDL 1011
Db      1052 AMWRPFGQLTTFVAFBARDPGRYLAIGDKFTRLPA-----YTSQDEKOLIRKLAPT 1104
QY      1012 VDAEYLVLPQCGFFCPDPAPGAGGMVHHRSSSTRSGGDLTLGLEPSEERAP----- 1065
Db      1105 TDGSEAIAPDDYLPKALGPS-----HRTDCT-----DEMPKLNRYC 1143
QY      1066 RPLAPSEGAGSDVFG----DLGMAAKGLQSLPTHDPQLQRYSDPTVPLPSETDGYV 1122
Db      1144 KPSNKNSTGDDERSSAREVGVGNLR-----LDLPVDEDDYL 1182
QY      1123 APLTCSPOEYVNPQDVRQPPSPREGPLPAARPAGATLERAKTLSPGKGVVVDVFAFG 1182
Db      1183 MP-TQPGNNNNNMN-----NPNQNNAAVGAAGYM-----DLIGVP 1220
QY      1183 GAVENPEYL---TPOGGAAPQPH-----PPPAFSP-AFDNLYYMD 1218
Db      1221 VSDNPEYLLNAQTGLVGSGEPIPTQIGIPVMGGPTMEVKVPMFGSEPTSSDHEYND 1279

RESULT 14
S35745
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus
C:Species: avian erythroblastosis virus
C:Date: 03-Mar-1994 #sequence_revision 26-May-1995 #text_change 28-Feb-1997
C:Accession: S35745
R:Vennstrom, B.
submitted to the EMBL Data Library, March 1993
A:Reference number: S35743
A:Accession: S35745
A:Molecule type: DNA
A:Residues: 1-544 <VEN>
A:Cross-references: EMBL:X12707
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific P
F:135-400/Domain: protein kinase homology <KIN>
F:143-151/Region: protein kinase ATP-binding motif
F:170/Active site: Lys #status predicted

Query Match      24.3%; Score 1647; DB 2; Length 544;
Best Local Similarity 54.9%; Pred. No. 1.4e-60;
Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15;

QY      572 GPEADQCVACAHYKDPFPFCVACPSGVKPDLSYMPWIKFPDDEGACQPCPINCTHSCVDL 631
Db      1 GP--DHCWKCAHFIDGPHCVKACPAVLGENDTL--VMKYADANAVCOLCHPNCTRGCKGP 57

QY      632 DDKGPAPQAFASPLTSIVSAVV-GILLVVVLGVVFGIILKRQOKIRKYTMRRLLQETEL 690
Db      58 GLEGCP---NGSKTPSIAAGVVGGLCLVAVGLIGLGLYLR--HIVKRTLRLLQEREL 113

```

```

QY      632 DDKGPAPQAFASPLTSIVSAVV-GILLVVVLGVVFGIILKRQOKIRKYTMRRLLQETEL 690
Db      58 GLEGCP---NGSKTPSIAAGVVGGLCLVAVGLIGLGLYLR--HIVKRTLRLLQEREL 113
QY      691 VPPLPSGAPNQAQMRILKETELRKVKVLSGSAFTGVYKGIWIPGENVKIPVAIKVLR 750
Db      114 VPPLPSGAPNQAHLRIILKETEFKVKVLGFGAFGVYKGLWIPGEKVTIPVAIKEUR 173
QY      751 ENTSPKANKEILDEAYVMAGVGSPPYVSRLLGLICLTSTVOLVTQMLPYGCLLDHVRNRR 810
Db      174 EATSPKANKEILDEAYVMASVDNPHVCRLLGLICLTSTVOLIITQMLPYGCLLDVIREHKON 233
QY      811 LGSQDLLNWCQIAKGMVLELDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETE 870
Db      234 IGSQYLLNWCQIAKGMVLEERHVMVHRDLAARNVLVKTPOHVKITDFGLAKOLGADEKE 293
QY      871 YHADGGKVPKIMWALESILRRFFTHQSDVWSGVTVWELMTFGAKPYDGPAREIPDLLE 930
Db      294 YHAGGKVPKIMWALESILHRYTHQSDVWSGVTVWELMTFGSKPYDGPASEISSVLE 353
QY      931 KGERLPQPPICITIDVYIMVVKWMDSECRPRFELVSEFSRMDPQRFVVIQ--NEDLG 989
Db      354 KGERLPQPPICITIDVYIMVVKWMSDASRPKFERELIAEFSKWARDPPRYLVITQGBRMH 413
QY      990 PASPLDSTFYRSLLEDDMDGLVDAEYLVPOOGFFCPDPAPGAGGMVHHRSSSTRSG 1049
Db      414 LPSPTDSKYRTLMBEEDMEDIVDAEYLVPHQGF-----NSPST--- 454
QY      1050 GGDLLTGLFSEEEAPRSL-----APSEGAGSDVFDGDLGMAAKGLQSLPTHDPSPLO 1104
Db      455 -----SRTPLSSLSATSNNSATNCIDRNG--H----- 481
QY      1105 RYSEDPVPLPSETDGYVAPLTCSPQEVVNPQDVRQPPSPREGPLPAARPAGAT--LER 1163
Db      482 -----PYREGDL-----PAPEYVQ--LMPKPPSTAMVQVQNIYVISTAIK 523
QY      1164 AKTLSPGKGVVVDVFAFGAVENPEYL 1191
Db      524 LPIDSRYQN-----SHSTAVDNPEYL 544

RESULT 15
S00727
kinase-related transforming protein (erbB) (EC 2.7.1.-) - avian erythroblastosis virus
C:Species: avian erythroblastosis virus
C:Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 28-Feb-1997
C:Accession: S00727
R:Scotting, P.; Vennstrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.
Oncogene Res. 1, 265-278, 1987
A:Title: Common site of mutation in the erbB gene of avian erythroblastosis virus mutat
A:Reference number: S00727; MUID:88217326; PMID:2897102
A:Molecule type: DNA
A:Residues: 1-545 <SCO>
A:Cross-references: EMBL:X06943
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; phosphotransferase
F:135-400/Domain: protein kinase homology <KIN>
F:143-151/Region: protein kinase ATP-binding motif

Query Match      24.2%; Score 1640; DB 2; Length 545;
Best Local Similarity 54.9%; Pred. No. 2.7e-60;
Matches 345; Conservative 69; Mismatches 122; Indels 92; Gaps 15;

QY      572 GPEADQCVACAHYKDPFPFCVACPSGVKPDLSYMPWIKFPDDEGACQPCPINCTHSCVDL 631
Db      1 GP--DHCWKCAHFIDGPHCVKACPAVLGENDTL--VMKYADANAVCOLCHPNCTRGCKGP 57

QY      632 DDKGPAPQAFASPLTSIVSAVV-GILLVVVLGVVFGIILKRQOKIRKYTMRRLLQETEL 690
Db      58 GLEGCP---NGSKTPSIAAGVVGGLCLVAVGLIGLGLYLR--HIVKRTLRLLQEREL 113

```



```

Qy 691 VEPLTPSGAMPNQMRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLR 750
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
114 VEPLTPSGEAPNQAHRLAIKETEFKVKVLGFGAFGVYKGLMIPGEKVTIPVAIKELR 173
Qy 751 ENTSPKANKIILDAYVMAGVSGPYVSRLLGICLTSTVQLVQIMPYGCLLDHVRENRR 810
Db      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
174 EATSPKANKIILDAYVMASVDNPHVCRLLGICLTSTVQLITQIMPYGCLLDVIREHKDN 233
Qy 811 LGSQDLNWCQIAKMGSYLEDVRLVHRDLAARNVLKSPNHVKITDPFGLARLLDDETE 870
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
234 IGSQYLLNWCQIAKGNVYLEERHLVHRDLAARNVLKTPQDVKITDPFGLAKQLGADEKE 293
Qy 871 YHADGGKVPKIKWMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLE 930
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
294 YHAEGGKVPKIKWMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIIPASEISSVLE 353
Qy 931 KGERLPOPPICTIDVYIMVCKWMIDSECRPRFRELVSFSPRMARDPQRFVVIQ-NEDLG 989
Db      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
354 KGERLPOPPICTIDVYIMVCKWMSDADSRPKFRELIAEFKWARDPPRYLVIQGDERMH 413
Qy 990 PASPLDSTFYRSLEDDMGDLVDAEYLYPQCGFFCFDPAPGAGGMVHRRHSSSTRSG 1049
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
414 LPSFTDSKFYRTLWEEEDMEDIVDAEYLYPHOGFF-----NSPST--- 454
Qy 1050 GGDLTLGLEPSEBEAPRSP-----APSEGAGSDVFDGDLGMAAGLQSLPTHDPSPLO 1104
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
455 -----SRTPLLSLSATSNNATNCIDRNG-----H----- 481
Qy 1105 RYSEDPVPLPSETDGVAPLTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAAGAT-LER 1163
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
482 -----PVREDGFL-----PAPEYVYVQ--LMPKPESTAMVQNIYNYISLTAISK 523
Qy 1164 AKTLSPGKNGVVKDVPFAFGGAVENPEYL 1191
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
524 LPMDSRYQN-----SHSTAVDNPEYL 544

```

Search completed: July 22, 2003, 09:07:45
Job time : 30.7967 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:25:54 ; Search time 22.9062 Seconds
(without alignments)
5267.077 Million cell updates/sec

Title: SEQ4-465-479-12
Perfect score: 6804
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGPTAENPEYGLDVPV 1255
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 73:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution..

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6631	97.5	1255	A24571	protein-tyrosine k
2	5851	86.0	1254	I48161	p-185 precursor -
3	5844	85.9	1260	TVRTNU	protein-tyrosine k
4	3121	45.9	1210	GQHUE	epidermal growth f
5	3087	45.4	1210	A53183	epidermal growth f
6	3074.5	45.2	1223	TVCHLV	epidermal growth f
7	2943.5	43.3	1308	A47253	epidermal growth f
8	2658	39.1	1166	S06142	protein-tyrosine k
9	2384.5	35.0	1342	A36223	kinase-related tra
10	2300.5	33.8	1339	JC4387	epidermal growth f
11	1766.5	26.0	698	TFVFLV	protein-tyrosine k
12	1703	25.0	604	TVTVUH	protein-tyrosine k
13	1647	24.2	544	S35745	protein-tyrosine k
14	1640	24.1	545	S00727	kinase-related tra
15	1623	23.9	540	B44776	protein-tyrosine k
16	1621	23.8	540	TVFEVB	protein-tyrosine k
17	1620.5	23.8	1330	GQFFE	epidermal growth f
18	1479	21.7	644	A36325	epidermal growth f
19	1292	19.0	1323	E88257	protein let-23 (im
20	1292	19.0	1374	S70712	protein-tyrosine k
21	1196	17.6	1369	S70713	protein-tyrosine k
22	1169	17.2	1717	A45558	epidermal growth f
23	1108	16.3	527	A42032	epidermal growth f
24	948.5	13.9	843	A27131	epidermal growth f
25	806.5	11.9	346	S13807	protein-tyrosine k
26	754.5	11.1	311	S13808	protein-tyrosine k
27	724	10.6	1363	T43220	insulin-like growt
28	696	10.2	1382	1 INUR	insulin receptor p
29	695	10.2	1383	2 A36080	insulin receptor p

RESULT 1

A24571

protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human

N:Alternate names: c-erb-B-2 protein precursor; Kinase-related transforming protein erbB

C:Species: Homo sapiens (man)

C>Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999

C:Accession: A24571; A25491; A44188; B44188; I59509; I57622

R:Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T.;

Nature 319, 230-234, 1986

A:Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth f

A:Reference number: A24571; MUID:86118663; PMID:3003577

A:Accession: A24571

A:Molecule type: mRNA

A:Residues: 1-1255 <YAM>

A:Cross-references: GB:U03363; NID:g31197; PIDN:CAA27060.1; PID:g31198

R:Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.

Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985

A:Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epiderm

A:Reference number: A25491; MUID:86016729; PMID:2995967

A:Accession: A25491

A:Molecule type: DNA

A:Residues: 737-1031 <SEM>

A:Cross-references: GB:M11767; NID:g182163; PIDN:AAA35808.1; PID:g553282

R:Coussens, L.; Yang-Peng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg, P.

Science 230, 1132-1139, 1985

A:Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chromos

A:Reference number: A44188; MUID:86070181; PMID:2999974

A:Accession: A44188

A:Molecule type: DNA

A:Residues: 740-910 <COUL>

A:Cross-references: GB:M12036; NID:g183988; PIDN:AAA35978.1; PID:g183989

A:Accession: B44188

A:Molecule type: mRNA

A:Residues: 1-517, 'RALL', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>

A:Cross-references: GB:M1730; NID:g183986

R:King, C.R.; Kraus, M.H.; Aaronson, S.A.

Science 229, 974-976, 1985

A:Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.

A:Reference number: I59509; MUID:85272597; PMID:2992089

A:Accession: I59509

A:Molecule type: DNA

A:Residues: 832-909 <REX>

A:Cross-references: GB:L29395; NID:g459807; PIDN:AAA35809.1; PID:g459808

R:Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.

Mol. Cell. Biol. 7, 2597-2601, 1987

A:Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptional

A:Reference number: I57622; MUID:87286898; PMID:3039351

A:Accession: I57622

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-191 <TAL>

A;Cross-references: GB:M16792; NID:G183983; PIDN:AAA58637.1; PID:G553332
 C;Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
 C;Genetics:
 A;Gene: GDB:ERBB2; NGL; HER-2
 A;Cross-references: GDB:120613; OMIM:164870
 A;Map position: 17q21.1-17q21.1
 A;Introns: 25/1; 75/3; 147/1; 883/3
 A;Note: the list of introns is incomplete
 C;Function:
 A;Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
 C;Superfamily: epidermal growth factor receptor; protein kinase homology
 C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
 inase
 F;1-21/Domain: signal sequence #status predicted <SIG>
 F;22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <WAT>
 F;22-653/Domain: extracellular #status predicted <EXT>
 F;70-304/Domain: EGF receptor extracellular domain repeat <EE1>
 F;395-605/Domain: EGF receptor extracellular domain repeat <EE2>
 F;654-675/Domain: transmembrane #status predicted <TM>
 F;676-983/Domain: intracellular #status predicted <INT>
 F;718-734/Domain: protein kinase homology <KIN>
 F;726-734/Region: protein kinase ATP-binding motif
 F;68,124,187,259,530,571,629/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F;686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
 F;753/Active site: Lys #status predicted
 F;1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 97.5%; Score 6631; DB 1; Length 1255;
 Best Local Similarity 97.6%; Pred. No. 1.3e-263;
 Matches 1225; Conservative 6; Mismatches 24; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLLALLPPGAASCTGCTDMKLRLLPASPETHLDMRLHYQCQVVGNNL 60
 DB 1 MELAALCRWGLLLALLPPGAASCTGCTDMKLRLLPASPETHLDMRLHYQCQVVGNNL 60
 QY 61 ELTYLPTNASLFLQDIQEVQVYLIAHNVQVPLRLIRVGTQLFEDNVALAVLNG 120
 DB 61 ELTYLPTNASLFLQDIQEVQVYLIAHNVQVPLRLIRVGTQLFEDNVALAVLNG 120
 QY 121 DPLNNTTPTVGASPGGLRELQRLSTEILKGGVLLQRPOLCVQDTILWKDIFHKNQOLA 180
 DB 121 DPLNNTTPTVGASPGGLRELQRLSTEILKGGVLLQRPOLCVQDTILWKDIFHKNQOLA 180
 QY 181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDQCSLTRVCAGGCARCKGPLTDCHEQC 240
 DB 181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDQCSLTRVCAGGCARCKGPLTDCHEQC 240
 QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
 DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
 QY 301 YNLYSTDVGSCTLVCPHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 360
 DB 301 YNLYSTDVGSCTLVCPHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 360
 QY 361 ELEFAGCKKIFGSLAFPLSPDGDASNTAPLQPELOVFETLEITGVLYISAMPDSL 420
 DB 361 IQEFAGCKKIFGSLAFPLSPDGDASNTAPLQPELOVFETLEITGVLYISAMPDSL 420
 QY 421 DLSVFQNLQVIRGRLHNGAYSLTLQGLGISWGLRSLRELGSGLALIHNNHLCFVHTV 480
 DB 421 DLSVFQNLQVIRGRLHNGAYSLTLQGLGISWGLRSLRELGSGLALIHNNHLCFVHTV 480
 QY 481 PNDQLFQVYIKANSKFIGITELECVGEGGLACHQLCARGHCWGPFGTQCNCQFIRGQEC 540
 DB 481 PNDQLFQVYIKANSKFIGITELECVGEGGLACHQLCARGHCWGPFGTQCNCQFIRGQEC 540
 QY 541 VVECRVLQGLPREYNARHCLPCHPECPONGSVTCFGEADQCVACAHYKDPPEFCVARC 600
 DB 541 VVECRVLQGLPREYNARHCLPCHPECPONGSVTCFGEADQCVACAHYKDPPEFCVARC 600
 QY 601 PSGVKPDLSPYMPFWKFPDEGACPCPINCTHSCVDLDDKGPAPQASPLTSIYSAVVG 660

DB 601 PSGVKPDLSPYMPFWKFPDEGACPCPINCTHSCVDLDDKGPAPQASPLTSIYSAVVG 660
 QY 661 ILLVVVLGVGILIKRQOKIRKYTMRLLOETVELVEPLTPSGAMPNQAOQRIKTEL 720
 DB 661 ILLVVVLGVGILIKRQOKIRKYTMRLLOETVELVEPLTPSGAMPNQAOQRIKTEL 720
 QY 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVRENTSPKANKEIIDEAYVMAGVSP 780
 DB 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVRENTSPKANKEIIDEAYVMAGVSP 780
 QY 781 YVSRLLGICLSTVOLVTQMLPYGCLLDHVRNRLGSDLLNMCQIAKMSYLEDVR 840
 DB 781 YVSRLLGICLSTVOLVTQMLPYGCLLDHVRNRLGSDLLNMCQIAKMSYLEDVR 840
 QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPDKWMALESILRRPT 900
 DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPDKWMALESILRRPT 900
 QY 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYMINVKCM 960
 DB 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYMINVKCM 960
 QY 961 IDSECRPRFRELVSFESFMSARDPQRFVVIQNEGLGPASPLDSTFYRSLLEDDMGDLVDA 1020
 DB 961 IDSECRPRFRELVSFESFMSARDPQRFVVIQNEGLGPASPLDSTFYRSLLEDDMGDLVDA 1020
 QY 1021 BEYLVPQGGFFCPDAPAGCAGGVHHRSSSTRSGGDLTLGLEPSEEAERPLAPSEG 1080
 DB 1021 BEYLVPQGGFFCPDAPAGCAGGVHHRSSSTRSGGDLTLGLEPSEEAERPLAPSEG 1080
 QY 1081 AGSDVDFDGLGWAAGKLOSLPTHDPSPLOQRYSDPTVPLPSETDGYVAPLTCSPQPEV 1140
 DB 1081 AGSDVDFDGLGWAAGKLOSLPTHDPSPLOQRYSDPTVPLPSETDGYVAPLTCSPQPEV 1140
 QY 1141 NQPDVRPOPSPRSGPLPAARPAAGATLERAKTLSPGKNGVVDVPAFGAVENPEYLTPO 1200
 DB 1141 NQPDVRPOPSPRSGPLPAARPAAGATLERAKTLSPGKNGVVDVPAFGAVENPEYLTPO 1200
 QY 1201 GGAAPOHPHPAPSPADNLYWDDPPERCAPSTFKGTPTAENPEYGLDVPV 1255
 DB 1201 GGAAPOHPHPAPSPADNLYWDDPPERCAPSTFKGTPTAENPEYGLDVPV 1255

RESULT 2
 I48161
 p-185 precursor - golden hamster
 C;Species: Mesocricetus auratus (golden hamster)
 C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999
 C;Accession: I48161
 R;Nakamura, T.; Ushijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishikawa,
 Gene 140, 251-255, 1994
 A;Title: Cloning and activation of the Syrian hamster neu proto-oncogene.
 A;Reference number: I48161; MUID:94193007; PMID:7908275
 A;Accession: I48161
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 1-1254 <RES>
 A;Cross-references: GB:D16295; NID:G493236; PIDN:BAA03801.1; PID:G747595
 C;Genetics:
 C;Superfamily: epidermal growth factor receptor; protein kinase homology
 C;Keywords: ATP
 F;718-983/Domain: protein kinase homology <KIN>
 F;726-734/Region: protein kinase ATP-binding motif

Query Match 86.0%; Score 5851.5; DB 2; Length 1254;
 Best Local Similarity 86.0%; Pred. No. 8.6e-232;
 Matches 1079; Conservative 61; Mismatches 114; Indels 1; Gaps 1;

QY 1 MELAALCRWGLLLALLPPGAASCTGCTDMKLRLLPASPETHLDMRLHYQCQVVGNNL 60
 DB 1 MELAALCRWGLLLALLPPGAASCTGCTDMKLRLLPASPETHLDMRLHYQCQVVGNNL 60

A:Molecule type: mRNA
A:Residues: 969-971, 'K', 973-1115, 'D' <EIS>
A:Cross-references: EMBL:Z12608
R:Heigermann, G.J.; Gill, G.N.
J. Biol. Chem. 263, 13152-13158, 1988
A:Title: Epidermal growth factor receptor threonine and serine residues phosphorylated i
A:Reference number: A28941; MUID:88330814; PMID:3138233
A:Accession: A28941
A:Molecule type: protein
A:Residues: 689-694, 'X', 696-704, 'L', 706-707, 989-992, 'XX', 995-996, 'X', 998-1000; 1002-1009,
R:Hibbs, M.L.; Dunn, A.R.; Alexander, W.S.
submitted to the EMBL Data Library, April 1994
A:Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor
A:Reference number: S45325
A:Accession: S45325
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-971, 'K', 973-1210 <VER>
A:Cross-references: EMBL:X78987; NID:g488830; PIDN:CAA55587.1; PID:g488831
R:Paria, B.C.; Das, S.K.; Andrews, G.K.; Dev, S.K.
Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993
A:Title: Expression of the epidermal growth factor receptor gene is regulated in mouse b
A:Reference number: I49643; MUID:93126380; PMID:7678348
A:Accession: I49643
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 12-20, 22-132 <RES>
A:Cross-references: GB:I06864; NID:g193001; PIDN:AAA53029.1; PID:g567201
C:Genetics:
A:Gene: EGFR
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphop
F:1-24/Domain: signal sequence #status predicted <SIG>
F:648-670/Domain: transmembrane #status predicted <TM>
F:712-977/Domain: protein kinase homology <KIN>
F:680,695/Binding site: phosphate (Thr) (covalent) #status experimental
F:720-728/Region: protein kinase ATP-binding motif
F:697,1070,1071/Binding site: phosphate (Ser) (covalent) #status experimental
F:993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental
F:1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental
F:1197/Binding site: phosphate (Tyr) (covalent) #status experimental
Query Match 45.4%; Score 3087; DB 2; Length 1210;
Best Local Similarity 49.5%; Pred. No. 5; e-119;
Matches 630; Conservative 168; Mismatches 363; Indels 112; Gaps 25;
QY 11 LLLALLPPGAA--STOVCTGTDMKRLPASPETHLMLRLHYOGCGVOVGNIJELTYLPTN 68
DB 14 LLTALCAAGALEEKVKCOQTSNRLTQLGTFEDHFLSLQRMVNNCEVILGNLEITYVQRN 73
QY 69 ASLSFLQDTQEVGYVLIHNOVROVPLQRLTRVGTQTFEDNYALAVLDNGDPLNNTTP 128
DB 74 YDLSFLKTTQEVAGYVLIATNTVERPLENQLIIRNLYENTYALAILSN----- 124
QY 129 VTGASPGGLRELQRLSLTBILKGVLIQRNPOLCYQDTILWKDI----FHKNNQLALILI 184
DB 125 -YGNRTGLRELPMRNLIQELLIGAVFNSNPILCNWDTIQWRDIVQNVFMSNMDL--- 180
QY 185 DYNRSRACHPCSPMKGSKCWGESSDDCSLRTVTCAGGCA-RCKGPLPTDCHEOCAAG 243
DB 181 -QSHPSSCPCKDPCSPGNGCWGGGEECNCKLTKIIICAOQCCHRCGRSPSDCCHNCAAG 239
QY 244 CTGPKHSDGLACLNHNSGICELHCPALVYNTDTFESMPNPEGRVTFGASCVTACPNY 303
DB 240 CTGPRESDCLVCKQFODEATCKDTCPLMLYNTTYQMDVNPBGKYSFGATCKKCPRY 299
QY 304 LSTDVGSCTLVCPLNHQVTAEDGTQRCBCKSKPCARVCYGLCMQVYKANSKFIGITELE 363
DB 300 VTDHGSVCYRACPDYIEV-EEDGIRKCKKCDGCKVCNGIGEFK-DTILSINATNTK 357
QY 364 -FAGCKKIFGSLAFLEPESFDGDPASNTAPLQPEQLQVFTLEITGYLVSANWPSLPDL 422
DB 358 HFKYCTAIGDHLILPVAFGKGSFTRTPPLDPRELEILTKVTEITGFLLIQAWPDNWTDL 417

RESULT 6
TVCHLV

epidermal growth factor receptor precursor - chicken

N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB

C:Species: Gallus gallus (chicken)

C>Date: 28-Feb-1985 #sequence_revision 05-May-1995 #text_change 04-Feb-2000

C:Accession: A27720; A00643

R:Jax, J.; Johnson, A.; Howk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Vennet

Mol. Cell. Biol. 8, 1970-1978, 1988

A:Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mo

A:Reference number: A27720; MUID:88261272; PMID:3260329

A:Accession: A27720
A:Molecule type: mRNA
A:Residues: 1-1223 <LAX>
A:Cross-references: GB:M20386
R:Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1985
A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A:Reference number: A00643; MUID:85228222; PMID:2988784
A:Accession: A00643
A:Molecule type: mRNA
A:Residues: 585-1223 <NIL>
A:Cross-references: GB:M10066
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
specific protein kinase
F:1-30/Domain: signal sequence #status predicted <SIG>
F:11-1223/Product: epidermal growth factor receptor #status predicted <MAT>
F:11-554/Domain: extracellular #status predicted <EXT>
F:81-307/Domain: EGF receptor extracellular domain repeat <EE1>
F:397-610/Domain: EGF receptor extracellular domain repeat <EE2>
F:655-677/Domain: transmembrane #status predicted <TM>
F:678-1223/Domain: intracellular #status predicted <INT>
F:719-984/Domain: intracellular #status predicted <KIN>
F:727-735/Region: protein kinase ATP-binding motif
F:136,202,280,361,370,422,575,580,615,635/Binding site: carboxydrate (Thr) (covalent) #
F:192,650/Binding site: carboxydrate (Ser) (covalent) #status predicted
F:687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:754/Active site: Lys #status predicted
F:1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat

Query Match 45.2%; Score 3074.5; DB 1; Length 1223;
Best Local Similarity 48.5%; Pred. No. 1.8e-118;
Matches 629; Conservative 173; Mismatches 349; Indels 147; Gaps 27;

QY 8 RGLLLALLPGAA-----STQVCTGDMKRLPASPETHDMLRHLVQGVQVQGNLE 61
DB 13 RGAALVLLVLLGVALCSAVEEKVCOGTNNKLTQLGHVEDHFTSLQRMNCEVLSNLE 72
QY 62 LLYLPTNASLSFLODIOEQGVYLAHNOVROVPLQRLIRVGTOLFEDNYALAVLDNGD 121
DB 73 IYVHNRLDTEFKTIOEVAGVFLALNMDVPIENQLIRGNVLYDMSFALAVLSNVH 132
QY 122 PLNNTPTVTVGASPGGLRELRLSLRSLTEILKGVLIQNPOLCYQDTILMKDIFHKNNQAL 181
DB 133 -NNKTQ-----GLRELPMKRLSEILNGVVKISNNPKLQNDIVLWNDIIDTSRK-PL 182
QY 182 TLID-TNRSRACHPCSPCKCKGRWCSESSEDQSLTRTVACGGCA-RCKGPLPTDCCHQ 239
DB 183 TVLDFASNLSSCPKCHPNCTEDHCWAGRGQNCQTLTKVCAQQCSGRCKGVPSDCCHQ 242
QY 240 CAAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGYTTCGASCVTAC 299
DB 243 CAAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGYTTCGASCVTAC 302
QY 300 PNYLSTDVGSCTVLCPLHNEVTAEDGTQRCCKSKPCARVCYGLGMOYIKANSKFYGI 359
DB 303 PHNVVTVHSGSVRSNCTDTEYV-BENGVRKCKCDGLCKSVKNGIGIGELAGILS-INA 360
QY 360 TELE-PAGCKKIFGSLAFIPESFDGDPASNTAPLEQLQVFTLEETIGYLYISAWPDS 418
DB 361 TNIDSFKNCTKINGDVSILPFAVLGDAFTKTLPLDPKLDVPTVKETISGFLLIQAWPDN 420
QY 419 LPDLSFQNLQVIRGILLHNGAYSLTQGLGLISWGLSLRELGSGLALIHNTLCPFH 478
DB 421 ATDLVAFENLEIIRTKOHQGYSLAVNMLKIQSLGLRSLKSIISGDITAMKNKLVAD 480
QY 479 TVPWDLQFQYIKANSKFYIGITECEVCGELACHOLCARGHCWGPQPTOCVNCOSFQ 538
DB 481 TNWNSLFAQSOKTKIIONRNKNDCTADRHVCDPLCSVDGCGWGPCHFCSCRFESRQK 540
QY 539 ECVECRVLQGLPREYVNRHCLPCHPECPQNG----SVTCFGPEADQCACAHYKDPFF 595

RESULT 7

A47253
epidermal growth factor receptor, HER4 - human
C:Species: Homo sapiens (man)
C:Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C:Accession: A47253
R:Plowman, G.D.; Culouscou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.; Ne;
Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993
A:Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epidermal
A:Reference number: A47253; MUID:93189574; PMID:8383326
A:Accession: A47253
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-1308 <PLO>
A:Cross-references: GB:L07868; NID:g337359; PIDN:AAB59446.1; PID:g337360
A:Note: sequence extracted from NCBI backbone (NCBIP:126842)
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; growth factor receptor
F:716-981/Domain: protein kinase homology <KIN>
F:724-732/Region: protein kinase ATP-binding motif

Db 541 ECVKOCNITQGEPRFEDSKLCHSECLVQNSTAYNTTCSGPGPDMCKAHFIDGPH 600
QY 596 CVARPCSGVKPDLSPYMPWKPEDEBGACQPCINCHSCVDLDDKGCPCAEQASPLTSTV 655
Db 601 CVKACPAVGENDTL-VWKYADANAVCOLCHPNCTRCCKGPGLEGCP---NGSKTPSTA 656
QY 656 SAVV-GILLVVVLGVFGILIKRROOKIRKYTMRLLOETELVELPTTPSGAMPNOAQMRI 714
Db 657 AGVGGLLCLVVGIGLGLYLRRL-HIVKRTLRLLQERELVEPLTPSGEAPNQAHLRI 715
QY 715 LKETELRKVKVLGSGAGFTVYKGIWIPGENVKIPVAIKVLRNTSPKANKIILDEAYVM 774
Db 716 LKETEFKVKVLGSGAGFTVYKGIWIPGEKVKIPVAIKELREATSPKANKIILDEAYVM 775
QY 775 AGVGSPPVYSLRLLIGTCTVTQVLTQLMYPYGLLDHVRNRRGLRSGDQLLNCWQIAKGMN 834
Db 776 ASVDNPHVCRLLGLTCTVTQVLTQLMYPYGLLDYIREHKONIGSQYLLNCWQIAKGMN 835
QY 835 YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESI 894
Db 836 YLEERLVRDLAARNVLKTPQHVKITDFGLAKLIGADEKEYHAEGGKVPKIMMALESI 895
QY 895 LRRRPTHOSDVMSYGVTVWELMTFGSKPYDGPASEISSVLEKGERLPQPPICITDVYMI 954
Db 896 LHRITTHOSDVMSYGVTVWELMTFGSKPYDGPASEISSVLEKGERLPQPPICITDVYMI 955
QY 955 MVKCMWIDSECRPRELVSFSSWARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDD 1013
Db 956 MVKCMWIDADSRPFRELIASFQKWARDPPRYLVIOGDERMHLPSFTOSKFTVRLMEED 1015
QY 1014 MGLVDAAEYLVPOQGFPCDPAPFAGAGMVHRRSSSTRSGGGDLTILGLEPSEEAPRS 1073
Db 1016 MEDIVDADEYLVPHQGF-----NSPST-----SRT 1041
QY 1074 PL-----APSGAGSDVFDGLGMAKGLQSLPTHDPSPLOKRYSEDPTVPLPST--DG 1126
Db 1042 PLLSLSLATSNSATNCID-----RNGQGHVPVREDSFQRYSSDPTGNFLEESIDDG 1093
QY 1127 YVAPLTCSPQPEYVNOQDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDYF- 1185
Db 1094 FL-----PAPEYVNO--LMPKFS-----TANQYQNIYNSL 1124
QY 1186 -----AFGAVENPEYLTPOGGAAPQHPHPPAFSPAFDNLVYDQ----- 1225
Db 1125 TAISKLPWDSYQNSHSTAVDNPEYL-----NTNOSPLAKTVFESSPYIQSGNHQI 1176
QY 1226 --DPPE-----RGAPPSTFKGTPTAENPEYLGLDVP 1254
Db 1177 NLDNPDYQODFLPNETKENGILLKVAENPEYLRVAAP 1214

Query Match		43.3%;	Score 2943.5;	DB 2;	Length 1308;
Best Local Similarity		45.2%;	Pred. No. 4.2e-113;		
Matches 610;		Conservative 181;	Mismatches 377;	Indels 183;	Gaps 30;
QY	9	WGILLALLPGAA-----STQCTGDMKRLPASPETHLDMLRHLVGGQVVOGNNLELT	64		
DB	8	WVSVSLVAAGTVQPSDSQSVCACTENKLSLSDLEQQYRALRKYKYEVEVVMGNLEITS	67		
QY	65	LPTNASLFLDIOEQVGVVLIAHNOVROVPLORLRIVRGTOLFDENYALAVLNDGDLN	124		
DB	68	IEHRDLSELRSVRETVGVVVALNQFRYLPLENLRIIRGTUVEDRYALALFLNVRKG	127		
QY	125	NTTPVTGASFGGLRELQLRSLTEILKGGVLIQNPOLCYQDTILMKDIFHKNQLALTII	184		
DB	128	NF-----GLQELGLKNTLEILNGVYVDQNKFLCYADTIHWQDIVRNPWPSNLTIV	178		
QY	185	DTNRSRACHPCSPMKGSRGWGSESDCQSLRTTVCAAGC-ARCKGPLEPTDCHEOCAAG	243		
DB	179	STNGSSGCGCRCHKSCGTG-RCWGPTEHNCOTLRTVCAEQCDGRCYGVYVSDCHRECAG	237		
QY	244	CTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPGRYTFGASCVTACPNNY	303		
DB	238	CSGPKDTCFACNFDNSGACVTCQCFVYNTPTTFOLEHNFNAKTYTGAFCKVCKPHNF	297		
QY	304	LSTDVGSCTLVCPHLNHQVTAEDGTQRCCKSKPCARVCYGLGMQVIKANSKFIGITELE	363		
DB	298	V-VDSSSCVACRSPSSKMEV-ENGIRKCKPCTDIPCACDGICTGSLMSAQTVDSNIDK	355		
QY	364	FAGCKIFGSLAPLPSFDGDPASNTAPIQPEOLQVFTELEITGYLXISAMPDLSLDLS	423		
DB	356	FINTCKINGNLPIVTHGIDPDYNAIEADPEKLNIVRTVREITGFLNQSPNNMTDFS	415		
QY	424	VFNQIVQIRGLIHNGAYSITLQGLGSLWGLRSRELGLALIHNTHLCFVHTVPWD	483		
DB	416	VFSNUTVIGRVLGSLILLKQQGITSLOFOSLKEISAGNIYITDNSLNCYHTYINTW	475		
QY	484	QLF-----ROYIKANSKFIGITELECVGEGGLACHOLCARGHCWGPGTQVCNCSQPLRQ	538		
DB	476	TLFSTINQIRIVIRDNK-----AENCTAEGMVCNHLCSDDGCGPGDQCLSCRRFSGR	530		
QY	539	ECVEBRVLQGLPREVYNARHCLPCHPEQP-QNGSVTCFGEADOCVCAHYKDPFPVC	597		
DB	531	ICIESCNLDGFEFEPENGISICVECDPQCEKMEGILLTCHGPGDNCCKSHFDKPGNCV	590		
QY	598	ARCPGSKPDLSPMPLTWKPDDEGACQPCINCTHSCVDLDDKGC-----PABQ	646		
DB	591	EKCPDLQGANF--IFKADPDRECHPCNPCTQCNGPTSHDCIYYPTWGHSTLPQHA	648		
QY	647	RASPLTSIVSAVV-GILLVVVLGVVFGILIKRQKIRKYTWRRLLIQTETELVEPLTPSGA	705		
DB	649	R-TPL--IAAGVIGGLFVLIVGLTFVAVVRRKSIK-KKRALRRL-ETELVEPLTPST	703		
QY	706	MPNQAMRILKETELRKVKVLGSGAGTVYKGIWIPDGENVKIPVAIKVLRNTPSPKAN	765		
DB	704	APNOAQLRLKETELRKRVVLGSGAGTVYKGIWPEGETVKIPVAIKILNETTGPKANV	763		
QY	766	EILDYAVNAGVSPVSRLLGICLTSTVOLTPMLPYCGLDHDVRENRGRIGSDLLNW	825		
DB	764	EFMDEALINWSDHPLHVLRLGVCUSPTIQLVTQLMPHGLLLEYVHEHKNIGSQLLNW	823		
QY	826	CMQIAKMSYLEDLVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVP	885		
DB	824	CVQIAKGMWYLBERRLVHRDLAARNVLKSPNHVKITDFGLARLLGDEGEKEYNADGGKP	883		
QY	886	IKWMALESILRRRPTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQP	945		
DB	884	IKWMALECIHYRKFTHQSDVWSYGVTVWELMTFGGKPYDGIPTREIPDLLEKGERLPQP	943		
QY	946	ICTIDVYIMVWCMWIDSCRRFRRELVSERFARMARDPQRFVYIONED-LGPASPLDSTF	1004		
DB	944	ICTIDVYVWVKWCMWIDADSRAPKFAAEFRMARDPQRYLVIQGDHMKLPSPNSDKF	1003		
QY	1005	YRSLLEDDMDGLVDABEYLVPQQGFPCPDPAFGAGGMVHRRHRSSTRSGGDLTLGLE	1064		

Db	1004	FOULLDEEDLEMDMAEYLVP-QAFNIPPP-----IYTSRAIDNSRS-----EIGHS	1051	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
----	------	--	------	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Db 123 YQK-NPSSP--DYVQVGLKQLQLSNLTSLSGGVKVSHPNPLNCLONVETINWWDIVDKTNP 179
QY 180 ALTLIDTNRSRACHPCSPCKGSRGWGESSEDCQSLTRTVACAGC-ARCKGLPLPTDCCHE 238
Db 180 TNNLIPHAFERQCKQCDHCVNGSCWAPGPGHCQKFTKLLCAEQCNRRCRGPKPIDCCNE 239
QY 239 QCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRTYTFGASCUTA 298
Db 240 HCAGCTGPRATDCLACRDFNDGCTKPPPKIYDIVSHQVVDNPNIKYTFGAACVKE 299
QY 299 CFYNYLSTDVSGCTLVCPLHNOEVTAEADGTQCEKSKPCARVCYGLGMOYIKANSKFTG 358
Db 300 CFESNVVTE-GACVRSNAGMLEVD-ENGKRCKPCDGVCPKVDGIGLSI-SMTIAVN 356
QY 359 ITEL-EFAGCKKIFGSLAFPLSPFDGPASNTAPLOEQLOVFETLEETIGLYLISAMPD 417
Db 357 STNIRSFNCTKINGDIILNRNSFEGDPHYKIGTMDPEHLNLTITVKEITGYLVIMWPE 416
QY 418 SLPDLVSFQNLQVIRGRIHNGAYS-LTLQGLGISWGLRSURELGSGLALHHTHLCF 476
Db 417 NMTLSVSFQNLLEIRGRTTFSGFVSFVVQVHRHQLGRLSKYSAGNVILKNTLQRLY 476
QY 477 VHTVPWDOLFROYIKANSKFIGITELCEVGEGLACHOLCARGHGWPGPTOCVNCQSLR 536
Db 477 ANTIWRLFRSEDSQ-----IYDARTENQTCNNESEDCGW-PGPTMCVSLHVD 528
QY 537 GQECVECKVLQGLPREYNARHCLPCHPECOFQNGSVTCFPEADQCACAHYKDPFPC 596
Db 529 GGRCVASCNLLGEPREAVQDGRVQCHQECIVQDLSLTCYGPANCKSAHFODGPOC 588
QY 597 VARCSPGVKPDLSYMPIKFPDEBACQPCPNCTHSCVDLDDKGPAPQASPLTSTIVS 656
Db 589 IPRCPHGILGDGDTL-INKYADKMGQCPCHQCTGCCGPGSLGCRGD-IVSHSSLAGV 646
QY 657 AVUGILLVVVLGVVFGILIKRQKIRKYTWERLLQETVELPELTPSGAMPNQAQWRLK 716
Db 647 LVSGLLTIVALLIIVLRRRIK-RKRTICLLQELVELBPLTPSGQAPNQAFLRLK 705
QY 717 ETELKRVKVLGSGAGTGYKGIWDGENVKIPVAIKVLRNTPSKANKEIIDEAYVMAG 776
Db 706 ETEFKKDRVLGSGAGTGYKGLWNPGENIRIPVAIKVLRNTPSKANKEIIDEAYVMAG 765
QY 777 VQSPVSRLLGLTSTVOLVQLMPYGLLDHVRNRRGLSGDQLLWNCQIAKMSYL 836
Db 766 VDPHVCRLLGLTCLSAVOLVQLMPYGLLDYVVRQHERICQGLLWNCVQIAKMSYL 825
QY 837 EDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIKMALESILR 896
Db 826 EERHLVHRDLAARNVLLKNPNHVKITDFGLSKLLTADEKEYQADGKVPKIKMALESILQ 885
QY 897 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMV 956
Db 886 WTYTHQSDVWSYGVTVWELMTFGSKPYDGIIPAKETASVLENGERLPQPPICTIEVYML 945
QY 957 KCWIDSECRPRFRELVSFSESMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDMDGD 1016
Db 946 KCWIDPSRPRFRELVSFSESMARDPGRYLIQ-----NLPSSLRLLFSRLSSDD--D 1000
QY 1017 LVDAEYLVPOQFCFPCPAPAGAGGVHRRSSSTRSGGDLTLGLPSEEEAPRSLA 1076
Db 1001 VDAEYLLPYKRI-----NRQGS-----EPCI 1023
QY 1077 PSEGAGSDVFDGLGWAAGKQLSLPDPSPLOQRYSEDPV-PLPSETDGYVAPLTCSP 1135
Db 1024 PPTGH-----PVRENSITLRSIDPTQNALEKLDGH----- 1055
QY 1136 QBEYVNPQDVRPQP-----PSPRE-----GPLP-AARPAGATLAKTSLPGKNGVVK 1182
Db 1056 --EYVNPQGETSSRLSDIYNPNYEDLTDGWPVSLSSQEAETNSRPEYLNNTQNSL-- 1111
QY 1183 DVFAEGGAVENPEYLTPOGGAAPQHPPPAPFADNLYYWDQDPPERGAAPPSTFKGPT 1242
Db 1112 -PLVSSGSMDDPDY---QAG-----YQAAF-----LPQTGALTNGMFLPA 1148

QY 1243 AENPEYLG 1250
Db 1149 AENLEYLG 1156
RESULT 9
A36223
kinase-related transforming protein (erbB3) (EC 2.7.1.-) precursor - human
C:Species: Homo sapiens (man)
C:Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000
C:Accession: A36223; I59164
R:Kraus, M.H.; Issing, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.
Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989
A:Title: Isolation and characterization of ERBB3, a third member of the ERBB/epidermal growth factor receptor gene family
A:Reference number: A36223; MUID:90083234; PMID:2687875
A:Accession: A36223
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1342 <KRA>
A:Cross-references: GB:M29366
R:Plowman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todaro, G.J.
Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990
A:Title: Molecular cloning and expression of another epidermal growth factor receptor-related gene
A:Reference number: I59164; MUID:90311312; PMID:2164210
A:Accession: I59164
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-559, 'G', 561-957, 'P', 959-1063, 'G', 1065-1342 <RES>
A:Cross-references: GB:M34309; NID:G183990; PIDN:AAA35979.1; PID:G306841
C:Genetics:
A:Gene: GDB:ERBB3; HER3
A:Cross-references: GDB:119880; OMIM:190151
A:Map position: 12q13-12q13
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homology
F:707-972/Domain: protein kinase homology <KIN>
F:715-723/Region: protein kinase ATP-binding motif
Query Match 35.0%; Score 2384.5; DB 2; Length 1342;
Best Local Similarity 40.0%; Pred. No. 2.8e-90;
Matches 529; Conservative 194; Mismatches 446; Indels 155; Gaps 35;
QY 10 GLLALALPPGAA--STQVCTGTDMLRLPASFPETHLMLRHLVQGCQVVOGNLELTLYLT 67
Db 11 GLLFLARGSEVGNSSQAVCPGTLNGLSVTGDAENQYQTLKLYERCEVVMGNLEIVLTGH 70
QY 68 NASLSFLDIDIEVOGYLIAHNQVRQVPLQRLIRVGTQLFEDNYVALVLDNGDPLNNTT 127
Db 71 NADLSFLQWIEZVTGYVLYVAMNEFTLPLNLRVVRGTQVYDGKFAIFVM-----LNNT 125
QY 128 PVTGASPGGLSELOLRSLTEILKGVLIQRPOLCYODTILWKDIFHKNNQLALTLIDTN 187
Db 126 ----NSSHALQQLKTLQTEILSGGVIEKNDKLCMDTIDWRDILVRDR-----AEIVVKD 178
QY 188 RSRACHPCSPCKGSRGWGESSEDCQSLTRTVACAGC-ARCKGLPLPTDCHEQCAAGCTG 246
Db 179 NGRSPPCHEVCKG-RCWGPGSEDCQTLTKTICAPQCNHCFGNPNQCCHDECAGGCG 237
QY 247 PKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRTYTFGASCVTACPVNYLST 306
Db 238 PQDTCFACRHFNDSGACVPRCPQPLVYVKNLTFLQLEPNPHTKYQYGVGVCAVSCPHNFV-V 296
QY 307 DVGSCITLVCPLHNOEVTAEADGTQCEKSKPCARVCYGLGMOYIKANSKF---IGITELE- 363
Db 297 DQTSVCRACPPDKMEVD-KNGLKMCPECGGLCPKACEGTG-----SGSRFQTVDSNIDG 350
QY 364 FAGCKKIFGSLAFPLSPFDGPASNTAPLOEQLOVFETLEETIGLYLISAWPDSLPLDS 423
Db 351 FVNCTKILGNLDLITGLTNGDPWHKIPALDEKLNVRFTVREITGYLNIQSWPHMNF 410
QY 424 VFQNLQVIRGRIHNGAYS-LTLQGLGISWGLRSURELGSGLALHHTHLCFVHTVPW 482

```

Db 411 VFSNLTTIGRSLYNRGFSLLIMKNLNVTSGLFRSLKEISAGRIYIISANRQLCYTHSLNW 470
QY 483 DOLFQRYIKANSKFITE-----LECVGEGGLACHOLCARGHCWGPGPTQVCNC 531
Db 471 TKVLR-----GPTERLDIKNRPDRCDVAGKVCDDPLSCGGCGGPGGQCLSC 520
QY 532 SOFLRQECVBCRVLQGLPREYVNNARHCLPCHPECPQNGSVTCFPGPADQVCAHAKY 591
Db 521 RNYSRGGVCTHCFNLGEPREFEAHEACFSCHPECPQMEGTATCNGSGSDTCAQCAHPR 580
QY 592 DPFFCVARCPGVKPDLSYMPWTWKFDPDEGACQPCPINCCHSCVDLDDKGCAPAEORA--- 648
Db 581 DDPHCVSSCPHGVLG--AKGPIYKYPDVQNECRPCHENCOTGCKGPELDCLGQTLVLIG 638
QY 649 -SPLTSIVAVVGILLVGVVFGILIKRRQKIR-KYTMERLLQETELVELPLTPSGAM 706
Db 639 KTHLTWALTIAAG--LVJIFMMLGGTFLYWRGRRIQNKRAMRYLGERGESIEPLDPS-BK 695
QY 707 PNAQMRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEI 766
Db 696 ANKVLARIFKETELRKLKVLGSGVFGTVHKGVWIPGESIKIPVCIKVIEDKSGRQSPA 755
QY 767 ILDEAYVMAGVSPYVSRLGLTSTVQLVTLQMPYGCILLDHVRENRLGSDLLNWC 826
Db 756 VTDHMLAIGSLDHAHIVRLGLCPGSSQLVTVQLPLGSLLDHVRQHRGALPQLLLNWG 815
QY 827 MOIAKMSYLEDLVRLVHRDLAARNVLKSPNVKLTDFGLRLDLDITEXHADGKVP 886
Db 816 VQIAKMYLEBGMVHRNLARNVLLKSPQVQVADFGVADLLPPDDKQLLYSEAKTPI 875
QY 887 KMALESILRRRFTHQSDVMSYGVTVWELMTCAKPYDGIPIAREIPDLLEKGERIPQPP 946
Db 876 KMALESIHFGKYTHQSDVMSYGVTVWELMTCAEPYAGLRLAEVPDLLEKGERLAQPI 935
QY 947 CTIDVYMWKCMWIDSECRPRELVSFERSMARDPQRFVVIQNEIDLGA---SPLDST 1003
Db 936 CTIDVYMWKCMWIDENIRPTEKELANEFTRMARDPPRYLVIKRES-GPGIAPGPEPHG 994
QY 1004 FYRSLLEDDMDGLVDABEYLVPOQGFCCPDAPGAGGMVHRHSSSTRSGGDLTLGL 1063
Db 995 LTNKLEVEVELPELDLDLDAEBD-----NLATTIGSALSULPV 1035
QY 1064 EP-SEEEAPRSLAPSEAGSDVFDGDLGMGAAGLQSLPTH--PSLQRYSEDTVPILP 1121
Db 1036 GTLNRPRGQSLLSPSSGY-MPMNQNLGSCQESAVSGSSERCPVSLH-----PMP 1088
QY 1122 -----SETDGVVA-----PLTCSPOPE-----YVNPQDVRPQPSPREG 1155
Db 1089 RGCLASESGEHTVGTSEAELOEKVSMCRSRSRSPRPRGDSAYHSQRHSLTTPVTPLSP 1148
QY 1156 P-----LPAARPAATLERAKTLP--GKNGV-----KDVFAFGAVENPEYL 1197
Db 1149 PGLEEDVNGYVMDTHLKTGTPSSREGTLSSVGLSVLGTBEDEB-----BEYEM 1200
QY 1198 TPQGAAPQPHPPAFSPAFDNLTYWD-----QDPPERGAPPSTFKGTPTAEN 1245
Db 1201 NRERRHSP-PHPPRSSLSEELGEYVMDVSGDLSASLGSTQSCPLHPVIMPTAGTTPDED 1259
QY 1246 PEYL 1249
Db 1260 YEYM 1263

```

RESULT 10

JC4387

epidermal growth factor receptor homolog precursor - rat

N.Alternate names: ErbB3 protein; HER3 protein

C.Species: Rattus norvegicus (Norway rat)

C.Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998

C.Accession: JC4387

R.Hellyer, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.

Gene 165, 279-284, 1995

A.Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protein.

A:Reference number: JC4387; MUID:96096535; PMID:8522190
 A:Accession: JC4387
 A:Molecule type: mRNA
 A:Residues: 1-1339 <HEL>
 A:Cross-references: GB:U29339; NID:9915389; PID:9915390
 A:Experimental source: liver
 A>Note: The authors translated the codon AAC for residue 369 as Thr and GTT for residue 370. This protein is a functional heregulin receptor that transduces signals to the cell.
 C:Genetics:
 A:Gene: ErbB3
 C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog
 C:Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane protein
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-1339/Product: epidermal growth factor homolog #status predicted <MAT>
 F:640-659/Domain: transmembrane #status predicted <TMM>
 F:705-970/Domain: protein kinase homology <KIN>
 F:713-721/Region: protein kinase ATP-binding motif
 F:939,1051,1156,1194,1196,1219,1257,1259,1273,1286,1325/Binding site: phosphate (Tyr) (c)

Query Match 33.8%; Score 2300.5; DB 2; Length 1339;
 Best Local Similarity 40.3%; Pred No. 7.5e-87;
 Matches 520; Conservative 173; Mismatches 429; Indels 167; Gaps 37;

```

QY 3 LAALCRWGLLLALLPPGAA---STQVCTGTDMKRLPASPEHLDMLRHLHYQGCVQGN 59
Db 7 LQVLC---FLLSLARGSEMGSQAVCPGTINGLSVTGDADNQYQTLVKLYEKCEVVMGN 62
QY 60 LELYPLPNASLSFLDIOEVQGVYLIHNOVROVPLORLIRIVRGTOQLFENYALAVLDN 119
Db 63 LEIVLTGHNADELQWIREVTAYVLVAMNEFSLPLPLNLRVVRGTQYVKGKFAIFVM-- 120
QY 120 GDPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQNPOLCYODTILWKDIFHKNNQL 179
Db 121 ---LNYNT---NASHALRQLKFTQLTEILSGGVYIEKNKDKLCHMDITDWRDVR--- 170
QY 180 ALTLDINRSRACHPCSPCMKSGRCWGESSDCQSLTRITVCAGG--ARKKGLPLPTDCHE 238
Db 171 GAEIVVKNNGANCPPEVCKG--RCWGPDPDCQILTKTICAPQCNRCFCFNPQNCCHD 229
QY 239 QCAAGCTGPKSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPERGTYFGASCUTA 298
Db 230 ECAGCGSPQDTCFACRRFNDGACVPRCEPLVYNNKLTFLQLEPNPHTKYQYGGVCVAS 289
QY 299 CPYNVSLTDVSGSTLVCPLHNOEVTAEQTCQKSKPCARVCVGL--GMQYTKANSKF 356
Db 290 CPHNFV-VDQTFVCVRCPPDRKVD--KHGLKMCPECGGLCPKACEGTGSGSRVQTVDSN 347
QY 357 IGITELEFAGCKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFEETLEEITGYLYISAWP 416
Db 348 ID---GFVNCTKILGNLDFLITGLNVDPWHKIIPALDPEKLVNVRTVREITGYLNIQSWP 403
QY 417 DSLPDLVSFQNLQVIRGRILHNGAYS-LTLOGLGISWGLSLRSLRELGLALIHHTHLC 475
Db 404 PHMNFVSFSLTITGRSLYNRGFSLLIMKNLNVTSGLFRSLKEISAGRIYIISANRQLC 463
QY 476 FVHTVPMDQLPR--QYTKANSKF---IGITELECVGEGGLACHOLCARGHCWGPGPTQCVN 530
Db 464 YHSLNLTWLRGSPSEERLDIKYDRPLG----ECLAEGKVCDDPLSCGGCGGPGAPQCLLS 519
QY 531 CSQFLRGQECVBEBCRVLQGLPREYVNNARHCLPCHPECPQNGSVTCFPGPADQVCACAHY 590
Db 520 CRNYSREGVCTHCFNLGEPREFEAHEACFSCHPECPQMEGTATCNGSGSDTCAQCAHPR 579
QY 591 KDPFFCVARCPGVKPDLSYMPWTWKFDPDEGACQPCPINCCHSCVDLDDKGCAPAEORA 648
Db 580 RDGPHCVNSCPHGILG--AKGPIYKYPDVQNECRPCHENCOTGCKGPELDCLGQTLVLIG 637
QY 649 SPLTSIVAVVGILLVGVVFGILIKRRQKIR-KYTMERLLQETELVELPLTPSGAMP 707
Db 638 SKPHLVIAVTVG--LAVILMILGGSFLYWRGRRIQNKRAMRYLGERGESIEPLDPS-BKA 694
QY 708 NQAQMRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEI 767

```

Db 695 NKVLARIFKTELRLKLVLSGVFGTVHKIGIWIPEGESIKIPVICIKVIEDKSGRSQFQAV 754
QY 768 LDEAYVMAGVSPYSVRLGICLTSTVQLVQVLTQMPYGCCLLDHVHNRGLSGQDLNWCW 827
Db 755 TDHMLAVGSLDHAHIVRLGLCPGSSQLVQVPLGLSLLDHVQKHRETLPQLLNWGV 814
QY 828 QIAKGMVSYLDELVRDLAARNVLKVSNNHYKIDTDFGLARLLDIDETEHADGKGVPIK 887
Db 815 QIAKGMVYLEHSMVHRDLAARNVLMKSPSQVQVADFGVADLLPPDKQLLHSEAKTPIK 874
QY 888 WMALESILRRRTHQSDVMSYGVTVWELMTFQAKPYDGIIPAREIPDLLEKGERLPPOPTIC 947
Db 875 WMALESIHFGKYTHQSDVMSYGVTVWELMTFQAEYPAGRLAEIPDLLEKGERLAQPOIC 934
QY 948 TIDVYMWKMWIDSECRPRFRELVSERFMRARDPQRFVVIQNEDELGPASPLDSTFYRS 1007
Db 935 TIDVYMWKMWIDENIRTFKELANETFRMARDPPRYLVIKRAS-GPGTP--PAAEPS 991
QY 1008 LLEDDMDGLVDAAEYLVPOQGFPCPDAPAGAGMVHRRHRSSTRSGGDLTLGLPSE 1067
Db 992 VLTTEL-----QEALEPEL-----DLDLEAE 1017
QY 1068 E-----EAPRSPLAPSEG-----AGSDVFDGLGMAAKGLQ 1099
Db 1018 EGLATSLGALSILPTGTLTRPGSQSLSPSSGYMWMQSSLGEACLDASVILGREGQFSR 1077
QY 1100 SLTPHDPSPLOYSEDPTVPLPSETDGVV-----APL-----TC-----SQPE-----Y 1139
Db 1078 PISLH-PIPRGR-----PASESEGHVGTGSEAEQEKVSVCRSRSRSPRGRDSAY 1129
QY 1140 VNQPDVRPOPSPREGP-----LPARPAGATLERAKTLSP-GKNGV----- 1181
Db 1130 HSQRHSLTPVPLPSPGLEEDGNGVYMPDTHLRGASSREGTLSSVGLSSVLGTEED 1189
QY 1182 KDVFAGGAVENPEYLTPOGGAAPOPHP 1210
Db 1190 ED-----EYEVNWRKRGSP-PRPP 1209
RESULT 11
TVFVLV
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian leukosis virus
N:Contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase
C:Species: avian leukosis virus, ALV
C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 11-Jun-1999
C:Accession: B00643; A00643
R:Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1985
A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A:Reference number: A00643; MUID:85228222; PMID:2988784
A:Accession: B00643
A:Molecule type: mRNA
A:Residues: 1-698 <NIL>
A:Cross-references: GB:M10066; GB:M13881; NID:g211749; PIDN:AAA48763.1; PID:g211750
A>Note: In Genbank entry CHKGRBBF, release 109.0, the source is designated as Gallus gal
C:Comment: This protein is synthesized as a gag-env-erbB protein.
C:Genetics:
A:Gene: gag-env-erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
F:1-6/Products: gag protein (fragment) #status predicted <GAG>
F:7-59/Product: env protein (fragment) #status predicted <ENV>
F:60-69/Product: protein-tyrosine kinase erbB #status predicted <ENB>
F:194-459/Domain: protein kinase homology <KIN>
F:202-210/Region: protein kinase ATP-binding motif
F:229/Active site: Lys #status predicted
Query Match 26.0%; Score 1766.5; DB 1; Length 698;
Best Local Similarity 52.2%; Pred. No. 2.5e-65;
Matches 374; Conservative 80; Mismatches 137; Indels 125; Gaps 18;
QY 578 GPEADQCVAHYKDPFPFCVACPSGVKPDLSYMPKPPDEGACQPCFNCTHSCVDL 637

Db 60 GP--DHCWKCHAFIDPHCHVCACPAGVLGENDTL-VWKYADANAVCQLCHPNCTGCKGP 116
QY 638 DDKGPABQBPASPLTSIVSAVV-GILLVVLGVWFGILIKRRQKIRKYTRMRLLQETEL 696
Db 117 GLEGCP--NSGKTPSIAAGVVGGLCLVWVGLGIGLYLRRR-HIVRKTRLRLLQEREL 172
QY 697 VEPLTPSGANPNQAMRILKETELRKVKVGLSGAPGTYYKGIWIPDGENVKIPVAKULR 756
Db 173 VEPLTPSGEAFNOAHLRIKETEFKVKVGLSGAFGTYYKGLWIPEGEKVKIPVAKELR 232
QY 757 ENTSPKANKEILDEAYVMAGVSPVSRLLGICLTSTVQLVQVLTQMPYGCCLLDHVHNRGR 816
Db 233 EATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQMLPYGCLLDYIIRHKON 292
QY 817 LGSQDLNWCQIAKGMVSYLDELVRDLAARNVLKVSNNHYKIDTDFGLARLLDIDETE 876
Db 293 IGSQVLLNWCQIAKGMVLEERRLVHRDLAARNVLVKTPOHVKITDFGLAKLILGADEKE 352
QY 877 YHADGGKVPILKWMALLESILRRFPTHQSDVMSYGVTVWELMTFQAKPYDGIIPAREIPDLLE 936
Db 353 YHAEGGKVPILKWMALLESILHRIYTHQSDVMSYGVTVWELMTFQSKPYDGIIPASEISSVLE 412
QY 937 KGERLUPQPICTIDVYMWKMWIDSECRPRFRELVSERFMRARDPQRFVVIQ-NEDELG 995
Db 413 KGERLUPQPICTIDVYMWKMWIDADSRRPRFRELIAEFKSNARDPPRYLVIOQDERMH 472
QY 996 PASPLDSTFYRSLLDEDDMDGLVDAAEYLVPOQGFPCPDAPAGAGMVHRRHRSSTRSG 1055
Db 473 LPSPTDSKFYRLMEEDMEDIVDADEYLVPHQGF-----NSFST--- 513
QY 1056 GGDLTGLPSEERAPRPL-----APSEGAGSDVFDGDLGMAAKGLQSLPTHDPSPILQ 1110
Db 514 -----SRTPLLSLSATSNNSATNCID-----RNCQGHVPREDVSFVQ 550
QY 1111 RYSEDPTVPLPSET--DCGVAPLTCSPQPEYVQNDVDPQPSREGPLPAARPAGATLE 1168
Db 551 RYSDPTGNFLEESIDDGFL-----PAPEYVQ--LMPKKPS----- 585
QY 1169 RAKTILSPGKNGVVKDVF-----AFGAVENPEYLTPOGGAAPQHPPPAF 1213
Db 586 -----TAMVQNIYNNISLTAISKLPMSRYQNSHSTAVDNPEYL-----NTNQSLA 633
QY 1214 SPAFDNLVYWDQ-----DPPE-----RGAPPSTFKGTPTAENPEYILGLDVP 1254
Db 634 KTVFESSPYWTQSGNHQINLDNPYQDPLPNETKPNGLLKVPAAENPEYILRVAAP 689
RESULT 12
TVYUHH
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain H)
C:Species: avian erythroblastosis virus
C:Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 11-Jun-1999
C:Accession: A00644; A38022
R:Ianamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.
Cell 35, 71-78, 1983
A:Title: The erbB gene of avian erythroblastosis virus is a member of the src gene famil
A:Reference number: A00644; MUID:84026539; PMID:6313229
A:Accession: A00644
A:Molecule type: DNA
A:Residues: 1-604 <YAM>
A:Cross-references: GB:K01216; NID:g209676; PIDN:AAA42400.1; PID:g209678
R:Debuire, B.; Henry, C.; Benatissa, M.; Bisette, G.; Claverie, J.M.; Saule, S.; Martin,
Science 224, 1456-1459, 1984
A:Title: Sequencing the erba gene of avian erythroblastosis virus reveals a new type of
A:Reference number: A38022; MUID:84223957; PMID:6328658
A:Accession: A38022
A:Molecule type: DNA
A:Residues: 1-28,'W',30-139,'F',141-145,'V',147-152 <DEB>
C:Cross-references: GB:K02006
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; oncogens; phosphotransferase; transforming protein; tyrosine-specific p

F;135-400/Domain: protein kinase homology <KIN>

RESULT 15
B44776
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain ES4)
C;Species: avian erythroblastosis virus
C;Date: 28-Apr-1993 #sequence_revision 28-Apr-1993 #text_change 04-Feb-2000
C;Accession: B44776
R;Bruskin, A.; Jackson, J.; Bishop, J.M.; McCarley, D.J.; Schatzman, R.C.
Oncogene 5, 15-24, 1990
A;Title: Six amino acids from the retroviral gene gag greatly enhance the transforming F
A;Reference number: A44776; MUID:90206603; PMID:1969616
A;Accession: B44776
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-540 <BRU>
A;Cross-references: GB:X52211
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; phosphotransferase; tyrosine-specific protein kinase
F;130-395/Domain: protein kinase homology <KIN>
F;138-146/Region: protein kinase ATP-binding motif

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:25:54 ; Search time 22.9062 Seconds
(without alignments)
5267.077 Million cell updates/sec

Title: SEQ4-579-593-12

Perfect score: 6803

Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVVP 1255

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 9613422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6651	97.8	1255	1 A24571	protein-tyrosine k
2	5856	86.1	1260	1 TVRTNU	protein-tyrosine k
3	5846.5	85.9	1254	2 I48161	p-185 precursor -
4	3116	45.8	1210	1 GQHUE	epidermal growth f
5	3089	45.4	1210	2 A53183	epidermal growth f
6	3064.5	45.0	1223	1 TVCHLV	epidermal growth f
7	2956.5	43.5	1308	2 A47253	epidermal growth f
8	2641	38.8	1166	1 S06142	protein-tyrosine k
9	2388.5	35.1	1342	2 A36223	kinase-related tra
10	2304.5	33.9	1339	2 J43387	epidermal growth f
11	1736.5	25.5	698	1 TVFVLV	protein-tyrosine k
12	1688	24.8	604	1 TVYUHV	protein-tyrosine k
13	1630.5	24.0	1330	1 G0PFFE	epidermal growth f
14	1617	23.8	544	2 S35745	protein-tyrosine k
15	1610	23.7	545	2 S00727	kinase-related tra
16	1593	23.4	540	2 B47776	protein-tyrosine k
17	1591	23.4	540	1 TVFVEB	protein-tyrosine k
18	1477	21.7	644	2 A36325	epidermal growth f
19	1294	19.0	1323	2 E88257	protein let-23 (im
20	1294	19.0	1374	2 S70712	protein-tyrosine k
21	1211	17.8	1369	2 S70713	protein-tyrosine k
22	1176	17.3	1717	1 A45558	epidermal growth f
23	1126	16.6	527	2 A20332	epidermal growth f
24	976.5	14.4	843	2 A27131	epidermal growth f
25	806.5	11.9	346	2 S13807	protein-tyrosine k
26	754.5	11.1	311	2 S13808	protein-tyrosine k
27	723	10.6	1363	2 T43220	insulin-like growth
28	697	10.2	1607	2 T43212	insulin-like growth
29	691	10.2	1383	2 A36080	insulin receptor p

RESULT 1

A24571

protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human

N;Alternate names: C-erb-B-2 protein precursor; kinase-related transforming protein erbB

C;Species: Homo sapiens (man)

C;Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999

C;Accession: A24571; A25491; A44188; B44188; I59509; I57622

R;Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T.;

Nature 319, 230-234, 1986

A;Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth f

A;Reference number: A24571; MUID:86118663; PMID:3003577

A;Accession: A24571

A;Molecule type: mRNA

A;Residues: 1-1255 <YAM>

A;Cross-references: GB:X03363; NID:g31197; PIDN:CAA27060.1; PID:g31198

R;Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.

Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985

A;Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epiderm

A;Reference number: A25491; MUID:86016729; PMID:2995967

A;Accession: A25491

A;Molecule type: DNA

A;Residues: 737-1031 <SEM>

A;Cross-references: GB:M11767; NID:gl82163; PIDN:AAA35808.1; PID:g553282

R;Cousens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg, P.

Science 230, 1132-1139, 1985

A;Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chromos

A;Reference number: A44188; MUID:86070181; PMID:2999974

A;Accession: A44188

A;Molecule type: DNA

A;Residues: 740-910 <COUL>

A;Cross-references: GB:M12036; NID:g183988; PIDN:AAA35978.1; PID:g183989

A;Accession: B44188

A;Molecule type: mRNA

A;Residues: 1-517; RALL, 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>

A;Cross-references: GB:M11730; NID:gl83986

R;King, C.R.; Kraus, M.H.; Aaronson, S.A.

Science 229, 974-976, 1985

A;Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.

A;Reference number: I59509; MUID:85272597; PMID:2992089

A;Accession: I59509

A;Status: translated from GB/EMBL/DBDJ

A;Molecule type: DNA

A;Residues: 832-909 <REX>

A;Cross-references: GB:I29395; NID:g459807; PIDN:AAA35809.1; PID:g459808

R;Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.

Mol. Cell. Biol. 7, 2597-2601, 1987

A;Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptional

A;Reference number: I57622; MUID:87286898; PMID:3039351

A;Accession: I57622

A;Status: translated from GB/EMBL/DBDJ

A;Molecule type: DNA

A;Residues: 1-191 <TAL>

A;Cross-references: GB:M16792; NID:g183983; PIDN:AAA58637.1; PID:g553332
C;Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
C;Genetics:
A;Gene: GDB:ERBB2; NGU; NEU; HER-2
A;Cross-references: GDB:120613; OMIM:164870
A;Map position: 17q21.1-17q21.1
A;Introns: 25/1; 75/3; 147/1; 883/3
A;Note: the list of introns is incomplete
C;Function:
A;Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
inase
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>
F;22-653/Domain: extracellular #status predicted <EXT>
F;70-304/Domain: EGF receptor extracellular domain repeat <EEI>
F;395-605/Domain: EGF receptor extracellular domain repeat <EE2>
F;654-675/Domain: transmembrane #status predicted <TM>
F;676-1255/Domain: intracellular #status predicted <INT>
F;718-983/Domain: intracellular kinase homology <KIN>
F;726-734/Region: protein kinase ATP-binding motif
F;68-124,187,530,571,629/Binding site: carboxylate (Asn) (covalent) #status predicted
F;686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F;753/Active site: Lys #status predicted
F;1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 97.8%; Score 6651; DB 1; Length 1255;
Best Local Similarity 97.6%; Pred. No. 2.1e-265;
Matches 1225; Conservative 10; Mismatches 20; Indels 0; Gaps 0;

```
QY 1 MELAALCRWGLLLALPPCASTVCTGDMKRLPASPETHLMLRHLRYGCGVQGNL 60
DB 1 MELAALCRWGLLLALPPCASTVCTGDMKRLPASPETHLMLRHLRYGCGVQGNL 60
QY 61 ELTYLPTNASLSFLQDIOEVQGVYLIHNNQVQLRQLRIVRGQLPEDNYVALVDNG 120
DB 61 ELTYLPTNASLSFLQDIOEVQGVYLIHNNQVQLRQLRIVRGQLPEDNYVALVDNG 120
QY 121 DPLNTTPTVGTASPGGLRELQRLSLEILKGVLTQRPOLCYQDTILKWKDIFHKNNQLA 180
DB 121 DPLNTTPTVGTASPGGLRELQRLSLEILKGVLTQRPOLCYQDTILKWKDIFHKNNQLA 180
QY 181 LTLIDNRSRACHPCSPCKSRGWSESDCOSLTRVTCAGGCARCKGLPTDCCHQC 240
DB 181 LTLIDNRSRACHPCSPCKSRGWSESDCOSLTRVTCAGGCARCKGLPTDCCHQC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRTYTFGASCVTACP 300
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRTYTFGASCVTACP 300
QY 301 NYNLSTDVGSCTLVCLNQNQVATBDGTQCEKSKPCARVCYGLGMQVIKANSKFIQIT 360
DB 301 NYNLSTDVGSCTLVCLNQNQVATBDGTQCEKSKPCARVCYGLGMQVIKANSKFIQIT 360
QY 361 ELEFAGCKKIFGSLAFLPESDGPASNTAPLOPELOVFEETLEITGVLYISAMPDSLP 420
DB 361 IOEFAGCKKIFGSLAFLPESDGPASNTAPLOPELOVFEETLEITGVLYISAMPDSLP 420
QY 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGSLWGLRSLRELGSGLALIHNNTHLCFVHTV 480
DB 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGSLWGLRSLRELGSGLALIHNNTHLCFVHTV 480
QY 481 PWDQLFRNPQHALLTANRPEDECVGEGLAHQLCARGHCWGPQTQCVCNSQFIRGQEC 540
DB 481 PWDQLFRNPQHALLTANRPEDECVGEGLAHQLCARGHCWGPQTQCVCNSQFIRGQEC 540
QY 541 VEECRVLQGLPREYNARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPPECVARC 600
DB 541 VEECRVLQGLPREYNARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPPECVARC 600
QY 601 PQYIKANSKFIQITELPDEEGACQPCINCTHSCVDLDDKGCAPAEQASPLTISIYAVVG 660
```

```
DB 601 PSGVKPDLSTYMPIWKFPEEGACQPCINCTHSCVDLDDKGCAPAEQASPLTISIYAVVG 660
QY 661 ILLVVVLGVVFGILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
DB 661 ILLVVVLGVVFGILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
QY 721 RKVKVLGSGAGFTGVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKELDEAYVMAGVGSF 780
DB 721 RKVKVLGSGAGFTGVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKELDEAYVMAGVGSF 780
QY 781 YVSRLLGLICTSTVQLVTQMLPYCCLLDHVRNRLGSLQDLNMCQIAKGMYSYLEDVR 840
DB 781 YVSRLLGLICTSTVQLVTQMLPYCCLLDHVRNRLGSLQDLNMCQIAKGMYSYLEDVR 840
QY 841 LVHRDLAARNVLKSPNHHVKITDFGLARLLDIDETEHADGGKVPICKMALESILRRRFT 900
DB 841 LVHRDLAARNVLKSPNHHVKITDFGLARLLDIDETEHADGGKVPICKMALESILRRRFT 900
QY 901 HQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLEKGERLPQPPICITIDVYMIWVKCWM 960
DB 901 HQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLEKGERLPQPPICITIDVYMIWVKCWM 960
QY 961 IDSECRPRFRELVSERFARMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDDGDLVDA 1020
DB 961 IDSECRPRFRELVSERFARMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDDGDLVDA 1020
QY 1021 BEYLVPQGGFCPPDPAAGAGMWHRRSSSTRSGGDLTLGLEPSEEEAPRSLPASEG 1080
DB 1021 BEYLVPQGGFCPPDPAAGAGMWHRRSSSTRSGGDLTLGLEPSEEEAPRSLPASEG 1080
QY 1081 AGSDVDFGDLGMAAGKLSLPTDPSPLQRYSDPTVPLPSETDGVVAPLTCSPQPEYV 1140
DB 1081 AGSDVDFGDLGMAAGKLSLPTDPSPLQRYSDPTVPLPSETDGVVAPLTCSPQPEYV 1140
QY 1141 NQPDVRFQPPSPREGPLPAARPAAGATLERAKTSLPGKNGVVKDYFAFGAVENPEYLTPO 1200
DB 1141 NQPDVRFQPPSPREGPLPAARPAAGATLERAKTSLPGKNGVVKDYFAFGAVENPEYLTPO 1200
QY 1201 GGAAPOHPHPAPSPAFDNLYYWQDPPERCAPPSTFGTPTAENPEYLGIDVPV 1255
DB 1201 GGAAPOHPHPAPSPAFDNLYYWQDPPERCAPPSTFGTPTAENPEYLGIDVPV 1255
```

RESULT 2
TVRTNU
protein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 11-Jun-1999
C;Accession: A24562; A61204
R;Bargmann, C.I.; Hung, M.C.; Weinberg, R.A.
Nature 319, 226-230, 1986
A;Title: The neu oncogene encodes an epidermal growth factor receptor-related protein.
A;Reference number: A24562; MUID:86118662; PMID:3945311
A;Accession: A24562
A;Molecule type: mRNA
A;Residues: 1-1260 <BAR>
A;Cross-references: EMBL:X03362; NID:g56745; PIDN:CAA27059.1; PID:g56746
R;Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohen,
Carcinogenesis 12, 1975-1978, 1991
A;Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals no
2-thiazolylformamide or N-methyl-N-nitrosourea.
A;Reference number: A61204; MUID:92035293; PMID:1682063
A;Accession: A61204
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 637-663,'V',665-702 <MAS>
A;Note: authors translated the codon GCA for residue 25 as Val
C;Genetics:
A;Gene: neu
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>

F:658-680/Domain: transmembrane #status predicted <TN>
F:723-988/Domain: protein kinase homology <KIN>
F:731-739/Region: protein kinase ATP-binding motif
F:711,191,263,535,576,634/Binding site: carbonyldehydrate (Asn) (covalent) #status predicted
F:691/Binding site: phosphate (Thr) (covalent) #status predicted
F:758/Active site: Lys #status predicted
F:1882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 86.1%; Score 5856; DB 1; Length 1260;
Best Local Similarity 86.0%; Pred. No. 7, 9e-233;
Matches 1081; Conservative 56; Mismatches 118; Indels 2; Gaps 2;

```
QY 1 MELAAALCRWGLLLALLPPGAASVCTGTDMKRLPASPETHDMLRHLVGGCQVVOGNL 60
DB 4 MELAAACRWGFLALLPPGIAGTQVCTGTDMKRLPASPETHDMLRHLVGGCQVVOGNL 63
QY 61 ELTYLPTNASLSFLQDIQEVQGYVLIHNOVROVQLRLRVGRQLFEDNYALAVLDNG 120
DB 64 ELTYVPANASLSFLQDIQEVQGYVLIHNOVROVQLRLRVGRQLFEDNYALAVLDNR 123
QY 121 DPLNNTTPTV-GASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILKWDIFHKNQOL 179
DB 124 DPQDNVAASVTPGRTPEGLRELQRLSLTEILKGGVLIQRPOLCYQDMVLAKDVFKNQOL 183
QY 180 ALTLIDTNRSRACHPCSPKCGSRGWGSSSDCQSLTRTVAGGCAKCGPLPTDCCHEQ 239
DB 184 APVDIDTNRSRACPPCAPACKDNHCWGESPEDCQILITCTTSGCARCKGRPLTDCCHEQ 243
QY 240 CAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTAC 299
DB 244 CAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTTC 303
QY 300 PNYLSTDVGSCTLVCPHNOEVTAEDGTQRCCKSKPCARVCYGLGMOYIKANSKFTGI 359
DB 304 PNYLSTEVGSCTLVCPHNOEVTAEDGTQRCCKSKPCARVCYGLGMEHLRGARITSD 363
QY 360 TELEFAGCKKIFGSLAFIPESFDGDPASNTAPLOPEQLQVFETLEETIGYLYISAWPDSL 419
DB 364 NVQEPDGCKKIFGSLAFIPESFDGDPSSGIAPLPEQLQVFETLEETIGYLYISAWPDSL 423
QY 420 PDLVSFONLQVIRGILHNGAYSILTOGLGTSWGLRLSLRELGLSGLALIHNTLHLCFVHT 479
DB 424 RDLVSFONLQVIRGILHNGAYSILTOGLGTHSLGRLSLRELGLSGLALIHNTLHLCFVHT 483
QY 480 VPWDLFRNPQALHTANRPEDE-CVGEGLACHOLCARGCWGPGTQCVNCSOFIRGQ 538
DB 484 VPWDLFRNPQALHTANRPEDELCVSSGLVCSLCAHGCWGPCTQCVNCSHFIRGQ 543
QY 539 ECVBEICRVQLGPREYVNAHCLPCHPECPQNGSVTCFGEADQCAAHYKDPFFCVA 598
DB 544 ECVBEICRVKGLPREYVSDKRLPCHPECPQNSSETCFGEADQCAAHYKDSSCVA 603
QY 599 RCPQVIKANSKFTIGTELPEDEGACQPCINCTHSCVDLDDKGPAPORASPLTSISAV 658
DB 604 RCPQSVKPDLSYMPITWKYPDEBEGICQPCINCTHSCVDLDDKGPAPORASPLTSISAV 663
QY 659 VGLILVVVLGVVFGILIKRRQKIRKYTMRLLOTELVEPLTPSGAMPNOAQRILKET 718
DB 664 EGVLLFLILVVVGLILIKRRQKIRKYTMRLLOTELVEPLTPSGAMPNOAQRILKET 723
QY 719 ELRKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAGV 778
DB 724 ELRKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAGV 783
QY 779 SPYVSRLIGICLTSTVOLVTOIMPYGCILLDHVRENHRLGSGODLLNWCQIAKMSYLE 838
DB 784 SPYVSRLIGICLTSTVOLVTOIMPYGCILLDHVRENHRLGSGODLLNWCQIAKMSYLE 843
QY 839 VRLHRLDAARNVLKSPNHVKITDFGLARLLDDDEYHADGGKVPITKMWALSIILRRR 898
DB 844 VRLHRLDAARNVLKSPNHVKITDFGLARLLDDDEYHADGGKVPITKMWALSIILRRR 903
QY 899 FTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIFDLLEKGERLPQPPTICTIDVYIMVYC 958
```

```
DB 904 FTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIFDLLEKGERLPQPPTICTIDVYIMVYC 963
QY 959 WMIDSECRPRFRELVSFSFARMARDPQRFVVIQNEIDLGPASPLDSTFFVRSLEDDDDMGDLV 1018
DB 964 WMIDSECRPRFRELVSFSFARMARDPQRFVVIQNEIDLGPASPLDSTFFVRSLEDDDDMGDLV 1023
QY 1019 DAEYLVLPQOGFFCPDPAPGAGGMVHRHRSSTRSGGDLTLGLEPSEEBAPRSLAPS 1078
DB 1024 DAEYLVLPQOGFFCPDPPTGPGTGAHRRHRSSTRSGGDLTLGLEPSEEBAPRSLAPS 1083
QY 1079 EGAGSDVFDGDLGNGAAKGLSLPHTDPSPLORYSEDPVPLPSETDGYVAPLTCSPQPE 1138
DB 1084 EGAGSDVFDGDLGNGAVTKGLQSLPHTDPSPLORYSEDPVPLPSETDGYVAPLTCSPQPE 1143
QY 1139 YVNOQDVRPQPPSPREGPLPAARPAAGATLERAKTSLPGKNGVVKDVFAGGAVENPEYLT 1198
DB 1144 YVNOQDVRPQPPSPREGPLPAARPAAGATLERAKTSLPGKNGVVKDVFAGGAVENPEYLT 1203
QY 1199 PQGGAAPQPPHPPAFSPAFDNLVYWDQDPPPERGAPPSTFKGTPTAENPEYLGDLVPV 1255
DB 1204 PREGTASPPHSPAFSPAFDNLVYWDQDPPPERGAPPSTFKGTPTAENPEYLGDLVPV 1260
```

RESULT 3

148161

p-185 precursor - golden hamster

C:Species: Mesocricetus auratus (golden hamster)

C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999

C:Accession: I48161

R:Nakamura, T.; Ushijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishikawa,

Gene 140, 251-255, 1994

A:Title: Cloning and activation of the Syrian hamster neu proto-oncogene.

A:Reference number: I48161; MUID:94193007; PMID:7908275

A:Accession: I48161

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1254 <RES>

A:Cross-references: GB:D16295; NID:G493236; PID:BAA03801.1; PID:G747595

C:Genetics:

A:Gene: neu

C:Superfamily: epidermal growth factor receptor; protein kinase homology

C:Keywords: ATP

F:718-983/Domain: protein kinase homology <KIN>

F:726-734/Region: protein kinase ATP-binding motif

Query Match 85.9%; Score 5846.5; DB 2; Length 1254;

Best Local Similarity 85.7%; Pred. No. 1.9e-232;

Matches 1075; Conservative 65; Mismatches 114; Indels 1; Gaps 1;

QY 1 MELAAALCRWGLLLALLPPGAASVCTGTDMKRLPASPETHDMLRHLVGGCQVVOGNL 60

DB 1 MELAAACRWGFLALLPPGIAGTQVCTGTDMKRLPASPETHDMLRHLVGGCQVVOGNL 60

QY 61 ELTYLPTNASLSFLQDIQEVQGYVLIHNOVROVQLRLRVGRQLFEDNYALAVLDNG 120

DB 61 ELTYVPANATLSFLQDIQEVQGYVLIHNOVROVQLRLRVGRQLFEDNYALAVLDNR 120

QY 121 DPLNNTTPTV-GASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILKWDIFHKNQOLA 180

DB 121 DPLDNVTATGRTPEGLRELQRLSLTEILKGGVLIQRPOLCYQDTILKWDIFHKNQOLA 180

QY 181 LTLDITNRSRACHPCSPKCGSRGWGSSSDCQSLTRTVAGGCAKCGPLPTDCCHEQC 240

DB 181 PVDIDTNRSRACPPCAPACKDNHCWGESPEDCQILITCTTGAAPRAVPAARALPTDCCHEQC 240

QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300

DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTTC 300

QY 301 YNYLSTDVGSCTLVCPHNOEVTAEDGTQRCCKSKPCARVCYGLGMOYIKANSKFTGIT 360

DB 301 YNYLSTEVGSCTLVCPHNOEVTAEDGTQRCCKSKPCARVCYGLGMEHLRGARITSAN 360

361 ELEFAGCKKIFGSLAFLPESPDGDPASNTAPLOPELOVFEETLEBITGYLYISAWPDSLP 420
Db IQEFAGCKKIFGSLAFLPESPDGDPSSGIAPLTPLOLQVFTEBITGYLYISAWPDSLH 420
QY DLSVFQNLQVIRGRILHNGAYSLTIQGLIGISWLGRLSRLRELSGLALIHNNHILCFVHTV 480
Db DLSVFQNLQVIRGRVLDHNGAYSLAQGLIGIRWLGRLSRLRELSGLVLIHRNTHILCFVHTV 480
QY PWDQLFRPHQALHTANRPEDECVGEGACHOLCARGHCWGPPTQCVNCSQFIRGQEC 540
Db PWDQLFRPHQALLHSGNPSSEECGLKDFACYPCLCAHGCWGPPTQCVNCSHFLRGQEC 540
QY VEECRVLQGLPREYVNAHCLPCHPECOPOGNSVTCFPEADQCVACAHYKDPPECVARC 600
Db VEECRVWGLPREYVNGHCLPCHPECOPOGNSVTCFSEADQCTACPHYKDSPPCVARC 600
QY POYIKANSKFITIGITLPEDEGACQPCINCTHSCVDLDDKGPAPQASPLTISIVAVVG 660
Db PSQVKPDLSYMPIWKYPDEEGWCQPCINCTHSCVDLDERGCPAEQASPLTISIATVVG 660
QY ILLVVVLGVFGILIKRQOKIRKYTMRLLOETELVPLTPSGAMPNOAQRILKETEL 720
Db ILLFLVIGVVGILIKRQOKIRKYTMRLLOETELVPLTPSGAMPNOAQRILKETEL 720
QY RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEIIDEAIVMAGVGP 780
Db RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEIIDEAIVMAGLGP 780
QY YVSRLLGICLTSTVQLVTLQMPYGLLDHVRNRLGRLSQDILLNMCQIAKGMYSLEDDR 840
Db YVSRLLGICLTSTVQLVTLQMPYGLLDHVRNRLGRLSQDILLNMCQIAKGMYSLEDDR 840
QY LVHRDLAARNVLKSPNVHKITDFGLARLLDIDETEHADGSKVPIKWALESILRRPT 900
Db LVHRDLAARNVLKSPNVHKITDFGLARLLDIDETEHADGSKVPIKWALESILRRPT 900
QY HQSDVWSYGVTVWELMTGAKPYDGI PAREIPDLLEKGERLPPOPICTIDVYIMVKCWM 960
Db HQSDVWSYGVTVWELMTGAKPYDGI PAREIPDLLEKGERLPPOPICTIDVYIMVKCWM 960
QY IDSECRPRFRELVSFSEMRARDPQRFVVIQNEBDLGPASPLDSTFYRSLLEDDMGDLVDA 1020
Db IDSECRPRFRELVSFSEMRARDPQRFVVIQNEBDLGPASPLDSTFYRSLLEDDMGDLVDA 1020
QY EBYLVPOQGFCDPAPGAGGMVHRHRSSTSRSGGLDLTIGLESBEAPRSLAPSEG 1080
Db EBYLVPOQGFCDPAPGAGGMVHRHRSSTSRSGGLDLTIGLESBEAPRSLAPSEG 1080
QY AGSDVFDGDLGMAAGLQSLPHTDPSPLORYSEDTPLPSETDGYVAPLTCSPQPEV 1140
Db AGSDVFEGLMGATKGPQISPRDLSPLORYSEDTPLPSETDGYVAPLTCSPQPEV 1140
QY NQPDVVRPOPSPREGPLPAARPAAGATLERAKTSLFGKNGVKDVFAGFAGAVENPEYLTPQ 1200
Db NQPEVRPOPPLTPREGPLPVRPAGATLERPKTSLFGKNGVKDVFTEGAVENPEYLVR 1200
QY GGAAQPHPPPPAFSPFNLYWQDPPRGAPPPTFGTPTAENPEYGLDVPV 1255
Db GGASQSPH-PPALCPAFNLYWQDPPRGSPFNTFECTPTAENPEYGLDVPV 1254

RESULT 4
GQHE
epidermal growth factor receptor precursor - human
N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
C:Species: Homo sapiens (man)
C:Date: 15-Nov-1984 #sequence revision 27-Nov-1985 #text change 11-Jun-1999
C:Accession: A00641; A25772; S30024; A38672; A00642; A43615; A23062; A05281; A50143; A33
R.Ullrich, A.; Coussens, L.; Hayflick, J.S.; Dull, T.J.; Gray, A.; Iam, A.W.; Lee, J.; I
rg, P.H.
Nature 309, 418-425, 1984
A:Title: Human epidermal growth factor receptor cDNA sequence and aberrant expression of

A:Reference number: A00641; MUID:84219729; PMID:6328312
A:Accession: A00641
A:Molecule type: mRNA
A:Residues: 1-1210 <ULL>
A:Cross-references: EMBL:X00588; NID:g31113; PIDN:CAA5240.1; PID:g757924
A:Note: the authors translated the codon AAG for residue 540 as Asn
R:Ishii, S.; Xu, Y.; Stratton, R.H.; Roe, B.A.; Merlino, G.T.; Pastan, I.
Proc. Natl. Acad. Sci. U.S.A. 82, 4920-4924, 1985
A:Title: Characterization and sequence of the promoter region of the human epidermal gro
A:Reference number: A25772; MUID:85270438; PMID:2991899
A:Accession: A25772
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-29 <ISH>
A:Cross-references: GB:M11234; NID:g181981; PIDN:AAA52370.1; PID:g553272
R:Haley, J.; Whittle, N.; Bennett, P.; Kinchington, D.; Ullrich, A.; Waterfield, M.
Oncogene Res. 1, 375-396, 1987
A:Title: The human EGF receptor gene: structure of the 110 kb locus and identification o
A:Reference number: S30024; MUID:88217333; PMID:3329716
A:Accession: S30024
A:Molecule type: DNA
A:Residues: 1-29 <HA2>
A:Cross-references: EMBL:X06370; NID:g31118; PIDN:CAA29668.1; PID:g31119
R:Haley, J.D.; Waterfield, M.D.
J. Biol. Chem. 266, 1746-1753, 1991
A:Title: Contributory effects of de Novo transcription and premature transcript terminat
A:Reference number: A38672; MUID:91107677; PMID:1988448
A:Accession: A38672
A:Molecule type: DNA
A:Residues: 1-29 <HA1>
A:Cross-references: GB:M38425; NID:g181977; PIDN:AAA63171.1; PID:g553271
A:Experimental source: carcinoma cell line A431-7
R:Xu, Y.; Ishii, S.; Clark, A.J.L.; Sullivan, M.; Wilson, R.K.; Ma, D.P.; Roe, B.A.; Mer
Nature 309, 806-810, 1984
A:Title: Human epidermal growth factor receptor cDNA is homologous to a variety of RNAs
A:Reference number: A00642; MUID:84245835; PMID:6330563
A:Accession: A00642
A:Molecule type: mRNA
A:Residues: 'RCATERRA', 150-187, 'KSVIQAV', 195, 'M', 197, 'A', 199-222, 'S', 224-304, 'RA', 307-321
, 798-799, 'TD', 802-811, 'R', 813-942 <XUY>
A:Experimental source: A431 human carcinoma cells, which have large numbers of EGF recep
R:Lin, C.R.; Chen, W.S.; Krueger, W.; Stolarsky, L.S.; Weber, W.; Evans, R.M.; Verma, I.
Science 224, 843-848, 1984
A:Title: Expression cloning of human EGF receptor complementary DNA: gene amplification
A:Reference number: A43615; MUID:84196372; PMID:6326261
A:Accession: A43615
A:Molecule type: mRNA
A:Residues: 713-964 <LIN>
A:Experimental source: epidermoid carcinoma cell line A431
R:Simmen, F.A.; Gope, M.L.; Schulz, T.Z.; Wright, D.A.; Carpenter, G.; O'Malley, B.W.
Biochem. Biophys. Res. Commun. 124, 125-132, 1984
A:Reference number: A23062; MUID:85046483; PMID:6093780
A:Accession: A23062
A:Molecule type: mRNA
A:Residues: 1028-1210 <SIM>
R:Weber, W.; Gill, G.N.; Speiss, J.
Science 224, 294-297, 1984
A:Reference number: A05281; MUID:84172183; PMID:6324343
A:Accession: A05281
A:Molecule type: protein
A:Residues: 25-30, 'S', 32-51, 454-467 <WEB>
R:Russo, M.W.; Lukas, T.J.; Cohen, S.; Staros, J.V.
J. Biol. Chem. 260, 5205-5208, 1985
A:Title: Identification of residues in the nucleotide binding site of the epidermal grow
A:Reference number: A60143; MUID:85182650; PMID:2985580
A:Accession: A60143
A:Molecule type: protein
A:Residues: 740-744, 'X', 746-747 <RUS>
R:Mroczkowski, B.; Mosig, G.; Cohen, S.
Nature 309, 270-273, 1984
A:Title: ATP-stimulated interaction between epidermal growth factor receptor and superco
A:Reference number: A38023; MUID:84191554; PMID:6325948
A:Contents: annotation; receptor activity

A:Molecule type: mRNA
 A:Residues: 969-971,'K',973-1115,'D' <EIS>
 A:Cross-references: EMBL:Z12608
 R:Heisermann, G.J.; Gill, G.N.
 J. Biol. Chem. 263, 13152-13158, 1988
 A:Title: Epidermal growth factor receptor threonine and serine residues phosphorylated in
 A:Reference number: A28941; MUID:88330814; PMID:3138233
 A:Accession: A28941
 A:Molecule type: protein
 A:Residues: 689-694,'X',696-704,'L',706-707;989-992,'XX',995-996,'X',998-1000;1002-1009,
 R:Hibbs, M.L.; Dunn, A.R.; Alexander, W.S.
 submitted to the EMBL Data Library, April 1994
 A:Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor
 A:Reference number: S45325
 A:Accession: S45325
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-971,'K',973-1210 <VER>
 A:Cross-references: EMBL:X78987; NID:G488830; PIDN:CAA55587.1; PID:G488831
 R:Paria, B.C.; Das, S.K.; Andrews, G.K.; Dey, S.K.
 Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993
 A:Title: Expression of the epidermal growth factor receptor gene is regulated in mouse b
 A:Reference number: I49643; MUID:93126380; PMID:7678348
 A:Accession: I49643
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 12-20,22-132 <RES>
 A:Cross-references: GB:L06864; NID:G193001; PIDN:AAA53029.1; PID:G567201
 C:Genetics:
 A:Gene: EGFR
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphop
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:648-670/Domain: transmembrane #status predicted <TM>
 F:712-977/Domain: protein kinase homology <KIN>
 F:720-728/Region: protein kinase ATP-binding motif
 F:680,695/Binding site: phosphate (Thr) (covalent) #status experimental
 F:697,1070/Binding site: phosphate (Ser) (covalent) #status experimental
 F:993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental
 F:1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental
 F:1197/Binding site: phosphate (Tyr) (covalent) #status experimental

Query Match 45.4%; Score 3089; DB 2; Length 1210;
 Best Local Similarity 49.6%; Pred. No. 1.6e-119;
 Matches 632; Conservative 165; Mismatches 364; Indels 112; Gaps 25;

Qy 11 LLLALLPGCAA--STOVCTGTDMLRLPASPTHLMDLRHLVYQGVQVQGNLELYLPTN 68
 Db 14 LTLALCAAGGALLEKKVCOGTSNRITQGTDFHFLSLQRYNCEVVLGNLEITYVQRN 73
 Qy 69 ASLSFLQDIQVQGYVLLAHNOVRQVQLRILVRGTQLFEDNYALAVLDNGDPLNNTTP 128
 Db 74 YDLSEFKTIQVAGVYVLLAHNTVERIPLENLQIRGNALYENTYALAILSN----- 124
 Qy 129 VTGASPGGLRELQRLSLTEILKGGVLIQBNPOLCYQDTILWKDI----FHKNQLALTLI 184
 Db 125 -YGTNRTGLRELPMENLOELIGAVRFNSNNPILCNDNTIQWRDVIQNVFMNSMDL--- 180
 Qy 185 DNNRSRACHPCSPMKGRCWSESSEDCOSLRTVCAGGCA-RCKGPLPTDCCHQCAAG 243
 Db 181 -OSHFSSCPKCDPSPNGSCWGGGNCQKLTIKIACQCSHRCGRSPSDCHNQCAAG 239
 Qy 244 CTGPKHSDCLAHFNHSGICELHCPALVTYNTDFESMPNPEGRTFGASCVTACPNY 303
 Db 240 CTGPRESCLVCKQFDEATCKDTCPLMLYNFTTYQMDVNPDEGKYSFGATCVKCKPRNY 299
 Qy 304 LSTDVGSCTLVCLPHNQVTAEDGTQRCCKSPCARVCYGLGMQYIKANSKFIGITILE 363
 Db 300 VYTDHGSCVRACGPDYIEV-EDDGIRKCKKCDPCRKVCNGIGIGEFK-DTLISINATNIK 357
 Qy 364 -FAGCKKIFGSLAFIPESFDGDPASNTAPLOPEQLQVTFLEITGYLYISAWPDSLPL 422
 Db 358 HFKYTAISGDHILHLPVAPKGSFTTTPPLDPRELEILKTVKEITGFLLIQAWPDNWDL 417

RESULT 6

TVCHLV

epidermal growth factor receptor precursor - chicken

N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB

C:Species: Gallus gallus (chicken)

C:Date: 28-Feb-1986 #sequence_revision 05-May-1995 #text_change 04-Feb-2000

C:Accession: A27720; A00643

R:Lax, I.; Johnson, A.; Howk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Vennet

Mol. Cell. Biol. 8, 1970-1978, 1988

A:Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mou

A:Reference number: A27720; MUID:88261272; PMID:3260329

A:Accession: A27720
A:Molecule type: mRNA
A:Residues: 1-1223 <LAX>
A:Cross-references: GB:M20386
R:Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1985
A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro-
A:Reference number: A00643; MUID:85228222; PMID:2988784
A:Accession: A00643
A:Molecule type: mRNA
A:Residues: 585-1223 <NIL>
A:Cross-references: GB:M10066
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
C:Specific protein kinase
E:1-30/Domain: signal sequence #status predicted <SIG>
F:31-1223/Product: epidermal growth factor receptor #status predicted <MAT>
F:31-654/Domain: extracellular #status predicted <EXT>
F:81-307/Domain: EGF receptor extracellular domain repeat <EE1>
F:397-610/Domain: EGF receptor extracellular domain repeat <EE2>
F:655-677/Domain: transmembrane #status predicted <TM>
F:678-1223/Domain: intracellular #status predicted <INT>
F:719-984/Domain: protein kinase homology <KIN>
F:727-735/Region: protein kinase ATP-binding motif
F:136,202,280,361,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent) #
F:192,650/Binding site: carbohydrate (Ser) (covalent) #status predicted
F:687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:754/Active site: Lys #status predicted
F:1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #sta-
Query Match 45.0%; Score 3064.5; DB 1; Length 1223;
Best Local Similarity 48.4%; Pred. No. 1.7e-118;
Matches 628; Conservative 173; Mismatches 350; Indels 147; Gaps 27;
Qy 8 RWG L L L L L P P G A A ----- S T V C T G T D M K L R L P A S P E T H D M L R H L Y G C Q V V Q G N L E 61
Db 13 R G A A V L L L L G V A L C S A V E E K V C Q G T N N K L T Q L G H V E D H F T S L Q R M Y N N C E V L S N L E 72
Qy 62 L T Y L P T W A S L S F L O D I Q E V G Y V L I A H N O V R Q P L Q R L R I V R G T Q L F E D N Y A L A V L N G D 121
Db 73 I T Y V E H R N R D L T F L K T I D E V A G Y L I A L M M V D V I P L E N L Q I R G N V L Y D N S F A L A V L S N Y H 132
Qy 122 P L N N T P T V T G A S P G L R E L O L R S L T E I L K G G V L F O R N P O L C Y O D T I L W K D I F H K N Q L A L 181
Db 133 - M N K T Q - - - - - G L R E L P M K R L S E I L N G V K V I N N P K L C N M D T V L W N D I I D T S R K - P L 182
Qy 182 T L I D - T N R S R A C H P C S P M K S R C W G S S E D C Q S L R T V C A G G C A - R C K G P L P T D C C H E Q 239
Db 183 T V L D F A S N L S S C P K C H P N C T H D C W G A G E Q N C Q T L T K V I C A Q Q C S G R C R G K V P S D C C H N Q 242
Qy 240 C A A C T G P K S D C L A C L H F N H S I G C E L H C P A L V T Y N T D T F E S M P N P E G R Y T F G A S C V T A C 299
Db 243 C A A C T G P R E S D C I A C R K F R D A T C K D T C P P L V N P T T Y Q M D V N P E G K Y S F G A T C V R E C 302
Qy 300 P Y N Y L S T D V G S C T L V C P L H N O E V T A E D G T O R C E K S C P C A R V C Y G L C M Q V I K A N S R F I G I 359
Db 303 P H N T V T D H G S C V R S C N T D T Y E V - E E N G V R K K C G L C S K V C N G I G I G E L K I L S - I N A 360
Qy 360 T E L E - F A G C K K I F G S L A F L P E S F D G D P A S N T A P L Q E Q L Q V F T L E I T G Y L I S A M P D S 418
Db 361 T N I D S F N C T K I N G D V S I L P V A F L G D A F T K T L P L D P K L D V F R T V K E I S F G L L I Q A W P D N 420
Qy 419 L P D L S V F Q N I Q V I R G R I I L H N G A V S L T I Q G L I S W L G R L S R L E I G S G L A L I H N T H L C F V H 478
Db 421 A T D L Y A F E N L E I I R G R T K Q H Q C Y S L A V V N L K I Q S L G L S R L K E I S D G I A I M K N K N L C Y A D 480
Qy 479 T V P W D Q L F R N P H O A L L H T A N R P E D C V G E G L A C H O L C A R G H C W G P G T C V C N S Q F L R Q 538
Db 481 T M N W R S L F A T Q S Q T K I I O N R N K N D C T A D R H V C D P L C S D V G C W G P G F H C F S C R F S R Q K 540
Qy 539 E C V E C R V L Q O L P R E Y N A R H C L P C H P E C O P Q N G - - - S V T C F G P E A D C C V A C A H Y K O P P E 595

QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRFT 900
 Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRFT 900
 QY 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMWKWM 960
 Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMWKWM 960
 QY 961 IDSECRPRRELVSFSEFMARDPQRFVVIQNEDLGPASPLDSTFYRSLLDDMDGLDVA 1020
 Db 961 IDSECRPRRELVSFSEFMARDPQRFVVIQNEDLGPASPLDSTFYRSLLDDMDGLDVA 1020
 QY 1021 EYVLVPOQGFCDPAPACAGVHHRSSSTRSGGDLTLGLEPSEEEAPRPLAPSEG 1080
 Db 1021 EYVLVPOQGFCDPAPACAGVHHRSSSTRSGGDLTLGLEPSEEEAPRPLAPSEG 1080
 QY 1081 AGSDVDFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
 Db 1081 AGSDVDFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
 QY 1141 NQPDVRRPQPSREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAPFGGAVENPEYLTPQ 1200
 Db 1141 NQPDVRRPQPSREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAPFGGAVENPEYLTPQ 1200
 QY 1201 GGAAPQHPPPAPSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255
 Db 1201 GGAAPQHPPPAPSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 6
 AAW92406
 ID AAW92406 standard; Protein; 1255 AA.
 AC AAW92406;
 XX
 XX 21-APR-1999 (first entry)
 DT
 XX Human HER-2/neu oncogene protein.
 DE
 XX HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;
 KW malignancy; treatment; tumour.
 KM
 XX Homo sapiens.
 OS
 XX
 XX Key Location/Qualifiers
 FH 676..1255
 FT Region
 FT /note= "region which elicits immune response"
 XX
 XX US5869445-A.
 XX
 XX 09-FEB-1999.
 XX
 XX 01-APR-1996; 96US-0625101.
 XX
 XX 01-APR-1996; 96US-0625101.
 PR 17-MAR-1993; 93US-0033644.
 PR 12-AUG-1993; 93US-0106112.
 PR 31-MAR-1995; 95US-0414417.
 XX
 XX (UNIW) UNIV WASHINGTON.
 PA
 XX
 XX Cheever MA, Disis ML;
 XX WPI; 1999-152835/13.
 DR N-PSDB; AAX01912.
 DR
 XX Use of HER-2/neu polypeptides - for eliciting an immune response to
 PT an HER-2/neu associated malignancy, particularly for treating or
 PT preventing tumours
 XX
 PS Claim 3; Column 31-38; 26pp; English.
 XX
 XX This sequence represents the human HER-2/neu oncogene protein. A fragment

CC of this protein is used in a method for eliciting or enhancing an immune
 CC response to HER-2/neu protein. The polypeptide can stimulate T cells and
 CC B cells to produce an immune response to the HER-2/neu protein. The
 CC method can be used for immunisation against a malignancy in which the
 CC HER-2/neu oncogene is associated and in the treatment of an existing
 CC tumour, or to prevent tumour occurrence or reoccurrence.
 XX
 SQ Sequence 1255 AA;
 Query Match 99.9%; Score 6806; DB 20; Length 1255;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1253; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MELAALCRNGLLLALLPPGAASSTQVCTGDMKRLRASPETHLDMLRHLYQGQVVGNNL 60
 Db 1 MELAALCRNGLLLALLPPGAASSTQVCTGDMKRLRASPETHLDMLRHLYQGQVVGNNL 60
 QY 61 ELTYLPTNASLSFLQDIOEVQGVLIHNOVQVPLQRLRIRVGTOLFEDNYALAVLDNG 120
 Db 61 ELTYLPTNASLSFLQDIOEVQGVLIHNOVQVPLQRLRIRVGTOLFEDNYALAVLDNG 120
 QY 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNNQLA 180
 Db 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNNQLA 180
 QY 181 LTLIDTNRSRACHPCSPMCKGSRMGESSEDCQSLTRTVACGCARCKGPLPTDCCHEQC 240
 Db 181 LTLIDTNRSRACHPCSPMCKGSRMGESSEDCQSLTRTVACGCARCKGPLPTDCCHEQC 240
 QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCYTACP 300
 Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCYTACP 300
 QY 301 YNYLSTDVGSCTLVCPLNHNOEVTAEQTCRCKSKPCARVCYGLGMEHLREVRAVTSAN 360
 Db 301 YNYLSTDVGSCTLVCPLNHNOEVTAEQTCRCKSKPCARVCYGLGMEHLREVRAVTSAN 360
 QY 361 IQEFAGCKKIFGSLAPLPSFDGDPASNTAPLQPEQLQVFETLEEITGYLISAWPDSL 420
 Db 361 IQEFAGCKKIFGSLAPLPSFDGDPASNTAPLQPEQLQVFETLEEITGYLISAWPDSL 420
 QY 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGISWGLRSLRELGLALIHNTLHLCFVHTV 480
 Db 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGISWGLRSLRELGLALIHNTLHLCFVHTV 480
 QY 481 PNDQLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPPTQCVCNCSQFLRGQEC 540
 Db 481 PNDQLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPPTQCVCNCSQFLRGQEC 540
 QY 541 VEECRVLOGLPREYVYNARHCLFCHPECPQNGSVTCFGEADQCVACAHYKDPFFCVARC 600
 Db 541 VEECRVLOGLPREYVYNARHCLFCHPECPQNGSVTCFGEADQCVACAHYKDPFFCVARC 600
 QY 601 PSGVKPDLSPYMPIWKPDPDEGACQPCINCTHSCVDLDDKGPAPORASPLTSIVSAVVG 660
 Db 601 PSGVKPDLSPYMPIWKPDPDEGACQPCINCTHSCVDLDDKGPAPORASPLTSIVSAVVG 660
 QY 661 ILLVVLGVVFGILIKRROOKIRKYTMRELLQETELVEPLTPSGAMPNQAQMRILKETEL 720
 Db 661 ILLVVLGVVFGILIKRROOKIRKYTMRELLQETELVEPLTPSGAMPNQAQMRILKETEL 720
 QY 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRRENTSPKANKEILDEYVMAGVGSF 780
 Db 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRRENTSPKANKEILDEYVMAGVGSF 780
 QY 781 YVSRLLGICLTSTVQLVTOLMPYGLLDHVRNRRGLSQDILLNWCMIKAGMSYLEYDR 840
 Db 781 YVSRLLGICLTSTVQLVTOLMPYGLLDHVRNRRGLSQDILLNWCMIKAGMSYLEYDR 840
 QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRFT 900
 Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRFT 900

Db 961 IDSECRFRRELVSERMRDPRFVVIQNEEDLGPASPLDSTFYRSLLEDMDGLVDA 1020
 Qy 1021 EYLVPOQGFCDPAPFAGAGMHHRRSSSTRSGGDLTLGLEPSEEEAPRSPAPSEG 1080
 Db 1021 EYLVPOQGFCDPAPFAGAGMHHRRSSSTRSGGDLTLGLEPSEEEAPRSPAPSEG 1080
 Qy 1081 AGSDVFGDILGMAAGLQSLPHDPSPLQRYSDPTVPLPSETDGVVAPLTCSPQPEYV 1140
 Db 1081 AGSDVFGDILGMAAGLQSLPHDPSPLQRYSDPTVPLPSETDGVVAPLTCSPQPEYV 1140
 Qy 1141 NOPDVRQPPSPREGPLPAARPAAGATLERAKTISPGKNGVVKOVAFEGGAVENPEYLTPQ 1200
 Db 1141 NOPDVRQPPSPREGPLPAARPAAGATLERAKTISPGKNGVVKOVAFEGGAVENPEYLTPQ 1200
 Qy 1201 GGAAPQHPHPAPSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGIDVPV 1255
 Db 1201 GGAAPQHPHPAPSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGIDVPV 1255

RESULT 8

AAY84780

ID AAY84780 standard; Protein; 1255 AA.

XX AAY84780;

DT 08-AUG-2000 (first entry)

XX Amino acid sequence of the SPLICE erbB-2 receptor protein.

XX SPLICE erbB-2 receptor protein; cell transformation disorder; cancer;
 KW tumor cell proliferation; tissue degeneration; arthropathy;
 KW bone resorption; inflammatory disease; degenerative disorder;
 KW wound healing.

XX Homo sapiens.

XX W0200020579-A1.

XX 13-APR-2000.

XX 01-OCT-1999; 99WO-CA00912.

XX 02-OCT-1998; 98US-0165192.

XX (UYMC-) UNIV MCMASTER.

XX Muller WJ, Siegel PM;

XX WPI; 2000-303768/26.

XX N-PSDB; AAA14812.

PT Nucleic acid encoding an erbB 2 receptor protein designated SPLICE
 PT erbB-2, inhibitors of the protein are useful for treatment of cancer -
 PS Claim 3; Fig 2; 60pp; English.

CC The present sequence represents a SPLICE erbB-2 receptor protein. The
 CC protein has an in-frame deletion of 16 amino acids, 2 of which are
 CC conserved cysteine residues, compared to the unspliced protein. The
 CC erbB-2 polynucleotide is used to construct probes for detecting
 CC disorders of cell transformation such as cancer. Antibodies to the
 CC protein may be used to detect SPLICE erbB-2 in a sample. Agents
 CC (e.g. antisense oligonucleotides) which inhibit the expression of
 CC SPLICE erbB-2 are useful for reducing tumor cell proliferation and
 CC treating cancer. Substances which stimulate SPLICE erbB-2 are useful
 CC for treating conditions involving damaged cells including conditions
 CC in which degeneration of tissue occurs, such as arthropathy, bone
 CC resorption, inflammatory diseases, degenerative disorders of the
 CC central nervous system and wound healing.

XX Sequence 1255 AA;

SQ

Query Match 99.9%; Score 6806; DB 21; Length 1255;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1253; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 MELAALCRWGLLLALLPPGAASQVCTGTDMLRLPASPEHLDMLRHLHYQGCVVQGNL 60
 Db 1 MELAALCRWGLLLALLPPGAASQVCTGTDMLRLPASPEHLDMLRHLHYQGCVVQGNL 60
 Qy 61 ELTYLPTNASLSFLQDIQEVQGVYLIHNRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
 Db 61 ELTYLPTNASLSFLQDIQEVQGVYLIHNRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
 Qy 121 DPLNNTPTVGTASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHNKQLA 180
 Db 121 DPLNNTPTVGTASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHNKQLA 180
 Qy 181 LTLIDTNRGRACHPCSPMCKGRCWGESSEDCQSLTRTVCAAGCARCKGPLPTDCCHEQC 240
 Db 181 LTLIDTNRGRACHPCSPMCKGRCWGESSEDCQSLTRTVCAAGCARCKGPLPTDCCHEQC 240
 Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 Qy 301 YNYLSTDVGSCTLVCPLNHNEVTAEDGTQRCCKSPCARVCYGLGMEHLREVRVTSAN 360
 Db 301 YNYLSTDVGSCTLVCPLNHNEVTAEDGTQRCCKSPCARVCYGLGMEHLREVRVTSAN 360
 Qy 361 IOEFACKIFGSLAPLPSFDGDPASNTAPLOPELOVFEETLEETIGVLYISAWPDSL 420
 Db 361 IOEFACKIFGSLAPLPSFDGDPASNTAPLOPELOVFEETLEETIGVLYISAWPDSL 420
 Qy 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGSIWGLRSLRELGSGLALIHNTLHLCFVHTV 480
 Db 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGSIWGLRSLRELGSGLALIHNTLHLCFVHTV 480
 Qy 481 PWDQLFRNPHQALLHTANRPEDECVGEGACHQLCARGHCWGPPTQCVNCSQFLRGQEC 540
 Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGACHQLCARGHCWGPPTQCVNCSQFLRGQEC 540
 Qy 541 VEECRVLQGLFREYNARHCLPCHPECPQNGSVTCFGEADQCVACAHYKPPFCVARC 600
 Db 541 VEECRVLQGLFREYNARHCLPCHPECPQNGSVTCFGEADQCVACAHYKPPFCVARC 600
 Qy 601 PSGVKPDLSPYMPKPPDEEGACQPCINCTHSCVDLDDKGCPEAQRASPLTSIVSAVVG 660
 Db 601 PSGVKPDLSPYMPKPPDEEGACQPCINCTHSCVDLDDKGCPEAQRASPLTSIVSAVVG 660
 Qy 661 ILLVVVLGVVFGILIKRROQKIRKVTMRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
 Db 661 ILLVVVLGVVFGILIKRROQKIRKVTMRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
 Qy 721 RKVKVLGSGAFGTVYKGIWIPGENVKIIPVAIKVLRENTSPKANKILDEAYMAGVGP 780
 Db 721 RKVKVLGSGAFGTVYKGIWIPGENVKIIPVAIKVLRENTSPKANKILDEAYMAGVGP 780
 Qy 781 YVSRLLGICLTSTVOLVTQIMPYGCLLDHVRNRRGLSGODLLNMCQIAKMSYLEVDYR 840
 Db 781 YVSRLLGICLTSTVOLVTQIMPYGCLLDHVRNRRGLSGODLLNMCQIAKMSYLEVDYR 840
 Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMWALESIILRRFT 900
 Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMWALESIILRRFT 900
 Qy 901 HQSDVWSYGVYTWELMTFGAKYDGIIPAREIPDLLEKGERLPOPPITCTIDVTMVKWM 960
 Db 901 HQSDVWSYGVYTWELMTFGAKYDGIIPAREIPDLLEKGERLPOPPITCTIDVTMVKWM 960
 Qy 961 IDSECRPRELVSERMRDPRFVVIQNEEDLGPASPLDSTFYRSLLEDMDGLVDA 1020
 Db 961 IDSECRPRELVSERMRDPRFVVIQNEEDLGPASPLDSTFYRSLLEDMDGLVDA 1020
 Qy 1021 EYLVPOQGFCDPAPFAGAGMHHRRSSSTRSGGDLTLGLEPSEEEAPRSPAPSEG 1080

Db	1021	EEYLVPPQGGFCPPAPACAGMVHRRSSSTRSGGDLTLGLEPSEEA	PRSLAPSEG	1080
QY	1081	AGSDVFDGDLGMAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV		1140
Db	1081	AGSDVFDGDLGMAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV		1140
QY	1141	NQDVRPQPPSPREGPLPAARPAAGATLERAKTLPSPGKNGVVVDVAFGGAVENPEYLTPO		1200
Db	1141	NQDVRPQPPSPREGPLPAARPAAGATLERAKTLPSPGKNGVVVDVAFGGAVENPEYLTPO		1200
QY	1201	GGAAQPHPPAPSPAFDNLVYWDQDPPRGAPSTFKGTPTAENPEYLGLDVVP		1255
Db	1201	GGAAQPHPPAPSPAFDNLVYWDQDPPRGAPSTFKGTPTAENPEYLGLDVVP		1255
RESULT 9				
AAB85458				
ID	AAB85458 standard; Protein; 1255 AA.			
XX	AAB85458;			
XX	AC			
DT	25-SEP-2001 (first entry)			
XX	Human HER-2/neu protein.			
XX				
KW	Antigen-presenting cell; immunogenic; immune response; HER-2/neu;			
KW	oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.			
OS	Homo sapiens.			
XX	WO200153463-A2.			
PN				
XX	26-JUL-2001.			
PD				
XX	19-JAN-2001; 2001WO-US01850.			
PF				
XX	21-JAN-2000; 2000US-0177545.			
XX	(CORI-) CORIXA CORP.			
XX	Cheever MA, Hand-Zimmermann S;			
PI	WPI; 2001-476112/51.			
XX	N-PSDB; AAH23392.			
DR				
XX	New antigen-presenting cells, useful as vaccines for eliciting or			
PT	enhancing an immune response to HER-2/neu protein, particularly useful			
PT	for treating or preventing cancer, e.g. breast cancer -			
XX	Claim 2; Page 41-46; 49pp; English.			
PS				
XX	The invention provides an isolated antigen-presenting cell, which			
CC	expresses at least an immunogenic portion of a polypeptide that produces			
CC	an immune response to HER-2/neu protein. The antigen-presenting cells are			
CC	useful as vaccines for eliciting or enhancing an immune response to			
CC	HER-2/neu protein, particularly in treating or preventing malignancies in			
CC	which the HER-2/neu oncogene is associated. Specifically, these are			
CC	useful for treating or preventing cancer, e.g. breast cancer, ovarian,			
CC	colon, lung or prostate cancers. The present sequence represents			
CC	the human HER-2/neu protein (also known as p185 or c-erbB2).			
XX	Sequence 1255 AA;			
SQ				
Query Match 99.9%; Score 6806; DB 22; Length 1255;				
Best Local Similarity 99.8%; Pred. No. 0;				
Matches 1253; Conservative 1; Mismatches 1; Indels 0; Gaps 0;				
QY	1	MELAALCRWGLLALLPPGAASQVCTGDMKLRLPASPTHLDMLRHLVYQGCQVQGNL	60	
Db	1	MELAALCRWGLLALLPPGAASQVCTGDMKLRLPASPTHLDMLRHLVYQGCQVQGNL	60	
QY	61	ELTYLPTNASLSFLQDIQEVGYVLIAHNQVRQVPLQRLIRVGTQLFEDNYALAVLDNG	120	

Db	61	ELTYLPTNASLSFLQDIQEVGYVLIAHNQVRQVPLQRLIRVGTQLFEDNYALAVLDNG	120
Qy	121	DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIIFHNKOLA	180
Db	121	DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIIFHNKOLA	180
Qy	181	LTLIDNRRACHPCSPCKGSCWGESSEDCOSLTRTVCAGGCARCKGKPLPTDCCHEQC	240
Db	181	LTLIDNRRACHPCSPCKGSCWGESSEDCOSLTRTVCAGGCARCKGKPLPTDCCHEQC	240
Qy	241	AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTACP	300
Db	241	AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTACP	300
Qy	301	YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSPCARVCYGLGMEHLREVRVTSAN	360
Db	301	YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSPCARVCYGLGMEHLREVRVTSAN	360
Qy	361	IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFTLEETITGYLYISAWPDSL	420
Db	361	IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFTLEETITGYLYISAWPDSL	420
Qy	421	DLVSFQNLQVIRGRILHNGAYSILTQGLGISMGLRSLRELGSGLALIHNTLHLCFVHTV	480
Db	421	DLVSFQNLQVIRGRILHNGAYSILTQGLGISMGLRSLRELGSGLALIHNTLHLCFVHTV	480
Qy	481	PMDQLFRNPHOALLHTANRPEDECVGEGLAGHQLCARGHGWGPGTQCVCNSQFLRGQEC	540
Db	481	PMDQLFRNPHOALLHTANRPEDECVGEGLAGHQLCARGHGWGPGTQCVCNSQFLRGQEC	540
Qy	541	VBECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFFCVARC	600
Db	541	VBECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFFCVARC	600
Qy	601	PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGPAPQASPLTISIVSAVVG	660
Db	601	PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGPAPQASPLTISIVSAVVG	660
Qy	661	ILLVVLGVVFGILIKRQOKIRKYTMRLLOQTELVEPLTPSGAMPNQAMRILKETEL	720
Db	661	ILLVVLGVVFGILIKRQOKIRKYTMRLLOQTELVEPLTPSGAMPNQAMRILKETEL	720
Qy	721	RKVVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPRANKILDEAYVMAGVSP	780
Db	721	RKVVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPRANKILDEAYVMAGVSP	780
Qy	781	YVSRLLGICLTSTVQLVTQMLPYGCLLDHVNRNRLGSLQDLNWCQIAKMSYLEDVR	840
Db	781	YVSRLLGICLTSTVQLVTQMLPYGCLLDHVNRNRLGSLQDLNWCQIAKMSYLEDVR	840
Qy	841	LVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKHWALESLRRRFT	900
Db	841	LVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKHWALESLRRRFT	900
Qy	901	HQSDVMSYGVYVWELMTFGAKPDVGIAREIPDLLEKGERLPPOPCTIDVYMWKCMW	960
Db	901	HQSDVMSYGVYVWELMTFGAKPDVGIAREIPDLLEKGERLPPOPCTIDVYMWKCMW	960
Qy	961	IDSECRPRFRELVSFSESRMARDPQRFVVIQNEEDLGPASPLDSTFYRSLLEDMDGLVDA	1020
Db	961	IDSECRPRFRELVSFSESRMARDPQRFVVIQNEEDLGPASPLDSTFYRSLLEDMDGLVDA	1020
Qy	1021	EYLVLPQGGFCPPDPAAGAGMVHRRSSSTRSGGDLTLGLEPSEEA PRSLAPSEG	1080
Db	1021	EYLVLPQGGFCPPDPAAGAGMVHRRSSSTRSGGDLTLGLEPSEEA PRSLAPSEG	1080
Qy	1081	AGSDVFDGLGMAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV	1140
Db	1081	AGSDVFDGLGMAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV	1140
Qy	1141	NQDVRPQPPSPREGPLPAARPAAGATLERAKTLPSPGKNGVVVDVAFGGAVENPEYLTPO	1200

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:25:54 ; Search time 22.9062 Seconds
(without alignments)
5267.077 Million cell updates/sec

Title: SEQ4-369-383-12
Perfect score: 6815
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGIDVPV 1255

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 73:*

1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6653	97.6	1255	1 A24571	protein-tyrosine k
2	5859	86.0	1260	1 TVRTNU	protein-tyrosine k
3	5951.5	85.9	1254	2 I48161	p-185 precursor -
4	3097	45.4	1210	1 GQHUE	epidermal growth f
5	3089	45.0	1210	2 A53183	epidermal growth f
6	3052.5	44.8	1223	1 TVCHLV	epidermal growth f
7	2919.5	42.8	1308	2 A47253	epidermal growth f
8	2636	38.7	1166	1 S06142	protein-tyrosine k
9	2366.5	34.7	1342	2 A36223	kinase-related tra
10	2280.5	33.5	1339	2 JC4387	epidermal growth f
11	1766.5	25.9	698	1 TVFVLV	protein-tyrosine k
12	1703	25.0	604	1 TVYUHV	protein-tyrosine k
13	1647	24.2	544	2 S35745	protein-tyrosine k
14	1640	24.1	545	2 S00727	protein-tyrosine k
15	1623	23.8	540	2 B44776	kinase-related tra
16	1621	23.8	540	1 TVFVEB	protein-tyrosine k
17	1600.5	23.5	1330	1 GQFPE	protein-tyrosine k
18	1461	21.4	644	2 A36325	epidermal growth f
19	1275	18.7	1323	2 E88257	epidermal growth f
20	1275	18.7	1374	2 S70712	protein let-23 [im
21	1192	17.5	1369	2 S70713	protein-tyrosine k
22	1161	17.0	1717	1 A45558	epidermal growth f
23	1084	15.9	527	2 A42032	epidermal growth f
24	945.5	13.9	843	2 A27131	epidermal growth f
25	806.5	11.8	346	2 S13807	epidermal growth f
26	754.5	11.1	311	2 S13808	protein-tyrosine k
27	730	10.7	1363	2 T43220	protein-tyrosine k
28	697	10.2	1382	1 INHUR	insulin-like growt
29	696	10.2	1383	2 A36080	insulin receptor p

insulin-like growt
insulin receptor p
protein-tyrosine k
insulin receptor-r
insulin receptor-r
insulin-like growt
insulin receptor -
insulin-like growt
insulin receptor (-
insulin receptor (-
protein-tyrosine k
protein-tyrosine k
protein-tyrosine k
protein-tyrosine k
tyrosine kinase Mp

ALIGNMENTS

RESULT 1

A24571
protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human
N:Alternate names: c-erb-B-2 protein precursor; kinase-related transforming protein erbB2
C:Species: Homo sapiens (man)
C:Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999
C:Accession: A24571; A24591; A44188; B44188; I59509; I57622
R:Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T.;
Nature 319, 230-234, 1986
A:Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth factor
A:Reference number: A24571; MUID:86118663; PMID:3003577
A:Accession: A24571
A:Molecule type: mRNA
A:Residues: 1-1255 <YAM>
A:Cross-References: GB:X03363; NID:g31197; PIDN:CAA27060.1; PID:g31198
R:Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.
Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985
A:Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epidermal growth factor receptor
A:Reference number: A25491; MUID:86016729; PMID:2995967
A:Accession: A25491
A:Molecule type: DNA
A:Residues: 737-1031 <SEM>
A:Cross-References: GB:M11767; NID:g182163; PIDN:AAA35808.1; PID:g553282
R:Coussens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg, P.
Science 230, 1132-1139, 1985
A:Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chromosomal location with c-erbB-2
A:Reference number: A44188; MUID:86070181; PMID:2999974
A:Accession: A44188
A:Molecule type: DNA
A:Residues: 740-910 <COU1>
A:Cross-References: GB:M12036; NID:g183988; PIDN:AAA35978.1; PID:g183989
A:Accession: B44188
A:Molecule type: mRNA
A:Residues: 1-517, 'RALL', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>
A:Cross-References: GB:M11730; NID:g183986
R:King, C.R.; Kraus, M.H.; Aaronson, S.A.
Science 229, 974-976, 1985
A:Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.
A:Reference number: I59509; MUID:85272597; PMID:2992089
A:Accession: I59509
A:Molecule type: DNA
A:Status: translated from GB/EMBL/DBJ
A:Residues: 832-909 <REX>
A:Cross-References: GB:L29395; NID:g459807; PIDN:AAA35809.1; PID:g459808
R:Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.
Mol. Cell. Biol. 7, 2597-2601, 1987
A:Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptional regulation
A:Reference number: I57622; MUID:87286698; PMID:3039351
A:Accession: I57622
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-191 <TAL>

A;Cross-references: GB:M16792; NID:G183983; PIDN:AAA58637.1; PID:G553332
C;Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
C;Genetics:
A;Gene: GDB:ERBB2; NGL; NEU; HER-2
A;Cross-references: GDB:120613; OMIM:164870
A;Map position: 17q21.1-17q21.1
A;Introns: 25/1; 75/3; 147/1; 883/3
A;Note: the list of introns is incomplete
C;Function:
A;Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
inase
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>
F;22-533/Domain: extracellular #status predicted <EXT>
F;70-304/Domain: EGF receptor extracellular domain repeat <BE1>
F;395-605/Domain: EGF receptor extracellular domain repeat <EE2>
F;654-675/Domain: transmembrane #status predicted <TM>
F;676-1255/Domain: intracellular #status predicted <INT>
F;718-983/Domain: protein kinase homology <KIN>
F;726-734/Region: protein kinase ATP-binding motif
F;68,124,187,259,530,571,629/Binding site: carboxydrate (Asn) (covalent) #status predict
F;686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F;753/Active site: Lys #status predicted
F;1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 97.6%; Score 6653; DB 1; Length 1255;
Best Local Similarity 97.6%; Pred. No. 6.5e-265;
Matches 1225; Conservative 8; Mismatches 22; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLALLPPGAASQVCTGTDMLRLPASPEHLDMLRLHYQGCVQGNL 60
DB 1 MELAALCRWGLLALLPPGAASQVCTGTDMLRLPASPEHLDMLRLHYQGCVQGNL 60
QY 61 ELTYLPTNASLFLQDIQEVQGVVLAHNOVQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLFLQDIQEVQGVVLAHNOVQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
QY 121 DPLNNTPTVTGSPGLRELQRLSLTEILKGGVLIQNPOLCYQDTILWKDIPHKNNOLA 180
DB 121 DPLNNTPTVTGSPGLRELQRLSLTEILKGGVLIQNPOLCYQDTILWKDIPHKNNOLA 180
QY 181 LTLDITNRSRACHPCSPMCKSGCWGESSEDCQSLRTVTCAGGCARCKGPLEPDDCCEQC 240
DB 181 LTLDITNRSRACHPCSPMCKSGCWGESSEDCQSLRTVTCAGGCARCKGPLEPDDCCEQC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPGRYTFFGASCVTACP 300
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPGRYTFFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPHNOEVTAEDGTQRCCKSPCARVCYGLGMQVIKANSKFIGHT 360
DB 301 YNYLSTDVGSCTLVCPHNOEVTAEDGTQRCCKSPCARVCYGLGMQVIKANSKFIGHT 360
QY 361 ELRFAGCKITFGSLAFIPESFDGDPASNTAPQVIKANSKFIGHTELTGVLITSAPDLSLP 420
DB 361 IQFAGCKITFGSLAFIPESFDGDPASNTAPLOEQLVFETUEETIGVLYISAWPDSLP 420
QY 421 DLSVFQNLQVIRGIRLLHNGAYSITLQGLGISWGLSLRELGSGLALIHNNHLCFVHTV 480
DB 421 DLSVFQNLQVIRGIRLLHNGAYSITLQGLGISWGLSLRELGSGLALIHNNHLCFVHTV 480
QY 481 PDQLFRNPHQALLHTANRPEDSCVGEGLACHQLCARGHCWGPGPTQVCNCSOFLRGQEC 540
DB 481 PDQLFRNPHQALLHTANRPEDSCVGEGLACHQLCARGHCWGPGPTQVCNCSOFLRGQEC 540
QY 541 VEBCRVLQGLPREYVVARHCLPCHBPCOPNGSVTCFGEADOCVCAHYKDPFPCVAC 600
DB 541 VEBCRVLQGLPREYVVARHCLPCHBPCOPNGSVTCFGEADOCVCAHYKDPFPCVAC 600
QY 601 PSQVKPDLSTYMPKIPFDEEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTSISAVVG 660
DB 601 PSQVKPDLSTYMPKIPFDEEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTSISAVVG 660

RESULT 2

TVRTNU

protein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 11-Jun-1999

C:Accession: A24562; A61204

R:Bargmann, C.I.; Hung, M.C.; Weinberg, R.A.

Nature 319, 226-230, 1986

A:Title: The neu oncogene encodes an epidermal growth factor receptor-related protein.

A:Reference number: A24562; MUID:86118662; PMID:3945311

A:Accession: A24562

A:Molecule type: mRNA

A:Residues: 1-1260 <BAR>

A:Cross-references: EMBL:X03362; NID:G56745; PIDN:CAA27059.1; PID:G56746

R:Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohen,

Carcinogenesis 12, 1975-1978, 1991

A:Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals no

2-thiazolyl]formamide or N-methyl-N-nitrosourea.

A:Reference number: A61204; MUID:92035293; PMID:1682063

A:Accession: A61204

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 637-663, 'Y', 665-702 <MAS>

A>Note: authors translated the codon GCA for residue 25 as Val

C:Genetics:

A:Gene: neu

C:Superfamily: epidermal growth factor receptor; protein kinase homology

C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosph

F;1-19/Domain: signal sequence #status predicted <SIG>

F;20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>

F:658-680/Domain: transmembrane #status predicted <TMN>
F:723-988/Domain: protein kinase homology <KIN>
F:731-739/Region: protein kinase ATP-binding motif
F:71,191,263,535,576,634/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:691/Binding site: phosphate (Thr) (covalent) #status predicted
F:758/Active site: Lys #status predicted
F:882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 86.0%; Score 5859; DB 1; Length 1260;
Best Local Similarity 86.0%; Pred. No. 1.8e-232;
Matches 1081; Conservative 54; Mismatches 120; Indels 2; Gaps 2;

QY 1 MELAALCRWGLLLALLPPGAASSTVCTGDMKRLRLPASPTHLDMLRHLYQGCVVQGNL 60
DB 4 MELAAWCRWGLLLALLPPGIAGTQVCTGDMKRLRLPASPTHLDMLRHLYQGCVVQGNL 63
QY 61 ELTYLPTNASLGFLODIQEVQGVLLIAHNVQVPLQRLIRVRGTQLFEDNVALAVLDNG 120
DB 64 ELTYVPANASLGFLODIQEVQGVLLIAHNVQVPLQRLIRVRGTQLFEDKVALAVLDNR 123
QY 121 DPLNNTPTVT-GASPGGLRELQRLSLEILKGGVLIQRNPOLCYQDTILWKDIFHKNQL 179
DB 124 DPQDNVAASTPGRTPEGLRELQRLSLEILKGGVLIQRNPOLCYQDMVLWKDVFRRKNQL 183
QY 180 ALTLIDTNSRACHPCSPMKSGRCSWGESSEDCQSILTRTVACGGCARCKGRLPTDCCHQ 239
DB 184 APVIDTNSRACPPCAPACKNHWCWESPEDCQILITGCTTSGCCARCKGRPLTDCCHQ 243
QY 240 CAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTAC 299
DB 244 CAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTTC 303
QY 300 PYNLSTVSGSCTLVCPFLHNQVETAEADGTQRCCKSKPCARVCYGLGMOYIKANSFIFI 359
DB 304 PYNLSTVSGSCTLVCPFPNNQVETAEADGTQRCCKSKPCARVCYGLGMEHLRGARITSD 363
QY 360 TELEFAGCKKIFGSLAFIPESPDGDPASNTAPQYIKANSKFIGITELTGLYLIASWPSDL 419
DB 364 NVQEFQCKKIFGSLAFIPESPDGDPSSGIAPLRPEQLQVFTEITGLYLIASWPSDL 423
QY 420 PDLVSFQNLQVLRILHNGVSLTQGLGISWGLRSRLGSLGIALIHNTHLCFVHT 479
DB 424 RDLVSFQNLQVLRILHNGVSLTQGLGISWGLRSRLGSLGIALIHNTHLCFVHT 483
QY 480 VPWQDLFRNPQALLHTANRPEDE-CVGBGLACHOLCARGHCWGPGPTQVCNCSQFLRQ 538
DB 484 VPWQDLFRNPQALLHSGNRPEEDLCVSSGLVCNLSCHGHCWGPGPTQVCNCSHFLRQ 543
QY 539 ECVEECRVLQGLPREVYNARHCLPCHPECPQNGSVTCFGEPAQCVAHAHYKDPFFCVA 598
DB 544 ECVEECRVWKGLPREVYSDKRLCPCHPECPQNSSETCFGEADQCAACAHYKDSSCVA 603
QY 599 RCPGSGVKPDLVYMPIWPKFDEGACOPCTNCTHSCVDLDDKGCRAEQASPLTSTVSAY 658
DB 604 RCPGSGVKPDLVYMPIWPKFDEEGICQPCPCTHSCVDLDERCPAEQASPVTFIATV 663
QY 659 VGILLVVVLGVFGIILKRRQKIRKYTWRRLLQETELVEPLTPSGAMPNQAOIRLKET 718
DB 664 EGVLPLFLVVVVGIIILKRRQKIRKYTWRRLLQETELVEPLTPSGAMPNQAOIRLKET 723
QY 719 ELRKVKVLGSGAPGVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVAVG 778
DB 724 ELRKVKVLGSGAPGVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVAVG 783
QY 779 SPVSVLLGICLTSTVOLVTLMPYGLLDHRENGRIGLSQDILLNKCQIAKMSYLE 838
DB 784 SPVSVLLGICLTSTVOLVTLMPYGLLDHREHGRIGLSQDILLNKCQIAKMSYLE 843
QY 839 VRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGKVPKIMMALESTLRR 898
DB 844 VRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGKVPKIMMALESTLRR 903
QY 899 FTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVKC 958

DB 904 FTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVKC 963
QY 959 WMIDSECRPRFRELVSERSMARDQRFVVIQNEDLGPASPLDSTFYRSLLEDDDDMGDLV 1018
DB 964 WMIDSECRPRFRELVSERSMARDQRFVVIQNEDLGPSSPMDSTFYRSLLEDDDDMGDLV 1023
QY 1019 DAEYLVQQQ3FFCPDPAPGAGVMVHRRHSRSTSGGGDLTLGLEPSEEEAPRSPAPS 1078
DB 1024 DAEYLVQQQ3FFSPDPPTGTSTARRHRSRSTSGGGELTLGLEPSEEGPPRSPAPS 1083
QY 1079 EGAGSDVDGDLGMAAKGLOSLPQRYSEDPTVPLPSETDGVVAPLTCSPQPE 1138
DB 1084 EGAGSDVDGDLGAMGVTKGLQSLSPQRYSEDTPLPPTDGTGVVAPLACSPQPE 1143
QY 1139 YYNQVDPVPPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKOVFAFGGAVENPEYLT 1198
DB 1144 YYNQSEVQPPPLTPEGPLPVPRPAGATLERPKTILSPGKNGVVKOVFAFGGAVENPEYLV 1203
QY 1199 POGGAAPQHPHPPAFSPADNLYYWDQPPPERGAPSTFKGPTTAENPEYLGLDVVP 1255
DB 1204 PREGTASPPHSPAFSPADNLYYWDQNSSEGGPPSPNPEGTPTAENPEYLGLDVVP 1260

RESULT 3
I48161
p-185 precursor - golden hamster
C:Species: Mesocricetus auratus (golden hamster)
C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999
C:Accession: I48161
R:Nakamura, T.; Ushijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishikawa, Y.
A:Title: Cloning and activation of the Syrian hamster neu proto-oncogene.
A:Reference number: I48161; MUID:94193007; PMID:7908275
A:Accession: I48161
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1254 <RES>
A:Cross-references: GB:D16295; NID:g493236; PIDN:BAA03801.1; PID:g747595
C:Genetics:
C:Gene: neu
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP
F:718-983/Domain: protein kinase homology <KIN>
F:726-734/Region: protein kinase ATP-binding motif

Query Match 85.9%; Score 5851.5; DB 2; Length 1254;
Best Local Similarity 85.7%; Pred. No. 3.7e-232;
Matches 1075; Conservative 64; Mismatches 115; Indels 1; Gaps 1;

QY 1 MELAALCRWGLLLALLPPGAASSTVCTGDMKRLRLPASPTHLDMLRHLYQGCVVQGNL 60
DB 1 MELAAWCRWGLLLALLSPGASGTQVCTGDMKRLRLPASPTHLDIVRHLYQGCVVQGNL 60
QY 61 ELTYLPTNASLGFLODIQEVQGVLLIAHNVQVPLQRLIRVRGTQLFEDNVALAVLDNG 120
DB 61 ELTYLPANATLSFLODIQEVQGVLLIAHNVQVPLQRLIRVRGTQLFEDKVALAVLDNR 120
QY 121 DPLNNTPTVTGASPGGLRELQRLSLEILKGGVLIQRNPOLCYQDTILWKDIFHKNQLA 180
DB 121 DPLDNVTATGTPTGRLRELQRLSLEILKGGVLIQRNPOLCYQDTILWKDIFRKNQLA 180
QY 181 LTLIDTNSRACHPCSPMKSGRCSWGESSEDCQSILTRTVACGGCARCKGRLPTDCCHQ 240
DB 181 PVIDITNSRACPPCAPACKNHWCWESPEDCQILITGIIAPRAVPAARARLPTDCCHQ 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTAC 300
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTTC 300
QY 301 YNVLSTVGSCTLVCPFLHNQVETAEADGTQRCCKSKPCARVCYGLGMOYIKANSFIFI 360
DB 301 YNVLSTVGSCTLVCPFLHNQVETAEADGTQRCCKSKPCARVCYGLGMEHLRGARITSA 360

A;Molecule type: mRNA
A;Residues: 969-971,'K',973-1115,'D' <EIS>
A;Cross-references: EMBL:Z12608
R;Heisermann, G.J.; Gill, G.N.
J. Biol. Chem. 263, 13152-13158, 1988
A;Title: Epidermal growth factor receptor threonine and serine residues phosphorylated
A;Reference number: A28941; MUID:98330814; PMID:3138233
A;Accession: A28941
A;Molecule type: protein
A;Residues: 689-694,'X',696-704,'L',706-707;989-992,'XX',995-996,'X',998-1000;1002-1009,
R;Hibbs, M.L.; Dunn, A.R.; Alexander, W.S.
submitted to the EMBL Data Library, April 1994
A;Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor
A;Reference number: S45325
A;Accession: S45325
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-971,'K',973-1210 <VR>
A;Cross-references: EMBL:X78987; NID:g488830; PIDN:CAA55587.1; PID:g488831
R;Paria, B.C.; Das, S.K.; Andrews, G.K.; Dev, S.K.
Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993
A;Title: Expression of the epidermal growth factor receptor gene is regulated in mouse
A;Reference number: I49643; MUID:93126380; PMID:7678348
A;Accession: I49643
A;Status: translated from GB/EMBL/DBEJ
A;Molecule type: mRNA
A;Residues: 12-20,22-132 <RES>
A;Cross-references: GB:L06864; NID:g193001; PIDN:AAA53029.1; PID:g567201
A;Genetics:
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphop
F;1-24/Domain: signal sequence #status predicted <SIG>
F;648-670/Domain: transmembrane #status predicted <TM>
F;712-977/Domain: protein kinase homology <KIN>
F;720-728/Region: protein kinase ATP-binding motif
F;680,695/Binding site: phosphate (Thr) (covalent) #status experimental
F;993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental
F;1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental
F;1197/Binding site: phosphate (Tyr) (covalent) #status experimental

Query Match 45.0%; Score 3069; DB 2; Length 1210;
Best Local Similarity 49.4%; Pred. No. 1.9e-118;
Matches 629; Conservative 165; Mismatches 367; Indels 112; Gaps 25;
QY 11 LLLALLPPGAA--STQVCTCTDMKRLPASPETHDMLRLHYOGCQVQVQGNLELTPLTN 69
DB 14 LTLALCAAGALBEKKVCQCTSNRLTQLGTFEDHFLSLQRMYNCEVVLGNLEITYVQRN 73
QY 69 ASLSFLQDIOEVQGYVLIHNOVROVPLQRLRIVRGTLQFEDNYALAVLDNGDPLNNTTP 128
DB 74 YDLSFLKTIQVAGYVLIANTVERIPLENLQIIRGNALYENTYALILSN----- 124
QY 129 VTGASPGGLRELQRLSITELKGVLTQRPOLCYODTILWKDI----FHKNQLALTILI 184
DB 125 -YGTNRITGLRELPMNLQELIIGAVRFSNPILCNMDTIQWRDIVQNVFMSNMSMDL--- 180
QY 185 DTRSRACHPCSPMKGSRGWGSESSDCQSLRTTVAGGCA-RCKGPLPTDCHQEOCAAG 243
DB 181 -QSHPSSCPCKDCSPGSCGWCWGEENCQKLTIICAQCCSHRCRGRSPSDCCHNQCAAG 239
QY 244 CTGPKHSDDLACLFHNSGICELHCPALVTYNTDTFESMENPEGRYTFGASCVTACPVNY 303
DB 240 CTGPRSDCLVCKQFQDEATCKDTCPPLMYNTTITQMDVNPBGKYSFGATCVKKCPRNY 299
QY 304 LSTDVGSCTLVCPHLNHQEVTAEDGTQRCERKSPCARVCYGLGMQYIKANSKFIGITELE 363
DB 300 VVTDHSGSVACGPDYVEV-EEDGIRKCKKCDGCRKVCNGIGIGEFK-DTLSINATNIK 357
QY 364 -FAGCKIFGSLAFLESFQDGPASNTAPQYIKANSKFIGITELTGLYVISAHPDSDPL 422
DB 358 FFKYCTAISGDLHLPLVAFKGSDFTRTPPLDPRLEILTKVKEITGFLLIQAWPDNWTDL 417

RESULT 6
TVCHLV

epidermal growth factor receptor precursor - chicken

N;Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB

C;Species: Gallus gallus (chicken)

C;Date: 28-Feb-1986 #sequence_revision 05-May-1995 #text_change 04-Feb-2000

C;Accession: A27720; A00643

R;Lax, I.; Johnson, A.; Hawk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Vennet

Mol. Cell. Biol. 8, 1970-1978, 1988

A;Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mo

A;Reference number: A27720; MUID:88261272; PMID:3260329

QY 423 SVFQNLQVIRGRILHNGAYSLTLOGLIGISWLGRLSRLRELGLALIHHTHLCFVHTVPW 482
DB 418 HAFENLEIRGRITKQHGQFSLAVGLNITSLGRLSKELSDGDVIISGNRLCYANTINW 477
QY 483 DOLFNRPHQALLHTANRPEDECVGEGLAHQICARGHCGWPGPTOCVNCQSLRQGEQVE 542
DB 478 KKLFGTGNQTKIMNRAEKCKAVNVHVCNPLCSSEGCGWGPEDPCVSCQNVSRGECVE 537
QY 543 ECRVLQGLPREVNVNARHCLPCHPECPQNGSVTCGPEADQCVACAHYKDPFCVACRPS 602
DB 538 KCNILEGEPRFVENSECICQHEPLCPQANNTCTGRGPDNCICQAHYIDGHVCTKCPA 597
QY 603 GYKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKCPAEQASPLTSTVSAVVGIL 662
DB 598 GIMGENNTL-VMKYADANNVCHLCHANCTYGCAGPLOGCEVWPSGPKIPSTATGIVGGL 656
QY 663 LVVVLGWVFEI-LIKRQOKIRKYTMRLLQTELVEPLTPSGAMPNOAMRILKETEURL 721
DB 657 LFTIV-VALGIGLFMRRRHIVKRTLRLRLQRELVEPLTPSGEAPNQAHLRIKETEFPK 715
QY 722 KUKVLGSGAFCTYKGINIPDGENVKIPVAIKVLRNTSPKANKELDEAYVMAGVGSY 781
DB 716 KIKVLGSGAFCTYKGLMIPGEKVKIPVAIKELREATSPKANKELDEAYVMASVDNPH 775
QY 782 VSRLLIGICLTSTVQLVTQLMPYGCILLDHVRENKRLGSDLLNWCQIAKMSYLEDVRL 841
DB 776 VCRLLIGICLTSTVQLITQLMPYGCILLDYVREKDNIGSQYLLNWCQIAKGNVLEDRRL 835
QY 842 VHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMKMALESILRRFFTH 901
DB 836 VHRDLAARNVLKTPQHVKITDFGLAKLIGAEKEYHAEGGKVPKMKMALESILHRIYTH 895
QY 902 QSDVMSYGYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPCTIDVYIMVKCWI 961
DB 896 QSDVMSYGYVWELMTFGSKPYDGIIPASDISSILEKGERLPQPPCTIDVYIMVKCWI 955
QY 962 DSECRPRFRELSEFSRMARDPQRFVIO-NEDLGPASPLDSTFYRSLLEDGMDLVDA 1020
DB 956 DADSRPKFRELLEFSQWARDQRYLVICQDERMHLPSFTDSNFYALMDEDMEDVDA 1015
QY 1021 EBYLVPOQGFPCPDPAAGAGGMVHRHRSSTRSGGDITLGLPSEESAPSPAPSG 1080
DB 1016 DEYLVPOQGF-----NSPST-----SRTPLLSLS 1041
QY 1081 AGSDVFDGDLGMAKGLQSLTHDPSPLORYSEDEPTVLPSET--DGVAELTCSQPE 1138
DB 1042 ATSN-----NSTVACINRNGSCRVKEDAFLOQYSSDPTGAVTEDNIDDAFL-----PVPE 1091
QY 1139 YVYQPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKGVVVKDVFAGGAVENPEYL- 1197
DB 1092 YVYQ-SVYKPRAGSVQNPVYHNPQLHP-----APGEDLHYQN--PHSNAVGNPEYL 1140
QY 1198 TPQGAAPQHPPPAFSPAFDNLYYDQ-----DP-----PERGAPSTFKGTP 1241
DB 1141 TAA-----PTCLSSGFNSPALWIKGSHOMSLDNDPDYQDFFPKETKNGIFKG-P 1190
QY 1242 TAENPEYLGLDVP 1254
DB 1191 TAENAEYLRVAPP 1203

A:Accession: A27720
A:Molecule type: mRNA
A:Residues: 1-1223 <LAX>
R:Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1985
A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A:Reference number: A00643; MUID:85228222; PMID:2988784
A:Accession: A00643
A:Molecule type: mRNA
A:Residues: 585-1223 <NIL>
A:Cross-references: GB:M10066
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
F:1-30/Domain: signal sequence #status predicted <SIG>
F:31-1223/Product: epidermal growth factor receptor #status predicted <MAT>
F:31-654/Domain: extracellular #status predicted <EXT>
F:81-307/Domain: EGF receptor extracellular domain repeat <EE1>
F:397-610/Domain: EGF receptor extracellular domain repeat <EE2>
F:655-677/Domain: transmembrane #status predicted <TM>
F:678-1223/Domain: intracellular #status predicted <INT>
F:719-984/Domain: protein kinase homology <KIN>
F:727-735/Region: protein kinase ATP-binding motif
F:136,202,280,361,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent) #s
F:192,650/Binding site: carbohydrate (Ser) (covalent) #status predicted
F:687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:754/Active site: Lys #status predicted
F:1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat
Query Match 44.8%; Score 3052.5; DB 1; Length 1223;
Best Local Similarity 48.3%; Pred. No. 9.1e-118;
Matches 627; Conservative 171; Mismatches 353; Indels 147; Gaps 27;
QY 8 RGLLLALLPPGAA-----STQVCTGDMKRLPASPETHLDMLRHLHYGCGQVQGNLE 61
DB 13 RGAALVLLVLLGVALCSAVEKKVCGQTNNKLTQLGHVEDHFTSLQRMYNCEVLSNLE 72
QY 62 LTVLPNASLSLQDIQEVGYVLIHNVQVPLQRLRIVRQTQIFEDNYALAVLDNGD 121
DB 73 ITVEHNRLDTLTKTQEVAGYVLIHNVQVPLQRLRIVRQTQIFEDNYALAVLSNYH 132
QY 122 PLNNTPTVTCASPGGLRELQLRLSLTEILKGGVLIORNPOLCYDTLLWKDIFHKNQAL 181
DB 133 -MNKTQ-----GLRELPMKLSLNGVVKISNNPKLNDMTVLNDDITSEK-PL 182
QY 182 TLID-TNRSRACHPCSPMKSGRCWGESSEDCSLRTRTVACGCA-RCXGPLPTDCCHQ 239
DB 183 TVLDFASNLSCPKCHPNCTEDHCWGAGQCQTLTKVCAQCQSGRCRGVPSDCCHQ 242
QY 240 CAACTGPKHSDCLACHFNHSGICELHCPALVYNTDTFESNPNEGRTFGASCVTAC 299
DB 243 CAACTGPKHSDCLACHFNHSGICELHCPALVYNTDTFESNPNEGRTFGASCVTAC 302
QY 300 PYNLYSTDVGSCTLVPLHNOETARDGTQRCCKSKPCARVCYGLGMVYKANSKFIGI 359
DB 303 PHNVVTDHSGVSRSCNTTYEV-FENGVRKCKDGLCKVCNGIGIGELKGILS-INA 360
QY 360 TELE-FAGCKKIFGSLAFIPESPDGPAANTAPQYIKANSKFTIGITELGYLYISAWPS 418
DB 361 TNIDSFNCKTNGDVSILPVAFLGDAFTKTLPLDPKLDVFTVKEISGFLLIQWPDN 420
QY 419 LPDLSVFONLQVIRGILHNGAYSLTLOGISWGLRSLRELGSGLAIHNNHLCFVH 478
DB 421 ATDLYAFENLEIIRGTRKHQGVSLAVNKLQSLGLRSLKESDGDIAIMKNKNCYAD 480
QY 479 TVPWDLFRNPHQALLHTANRPEDECVGSLACHOLCARGHCWGPGPTQVCNCSQFLRQ 538
DB 481 TNWRSLSFATQSKTKIQRNKNKNDCTADRHVCDPLCSDVGCWGCGPFHCFRCFRFSQK 540
QY 539 ECVEECRLVQLPREYVYNARHCLPCHPECPQNG---SVTCFGEADQCVACAHYKDPFF 595

RESULT 7

A47253
epidermal growth factor receptor, HER4 - human
C:Species: Homo sapiens (man)
C>Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C:Accession: A47253
R:Plowman, G.D.; Culouscou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.; N.
Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993
A:Title: Ligand-specific activation of HER4/pl180erbB4, a fourth member of the epidermal
A:Reference number: A47253; MUID:93189574; PMID:8383326
A:Accession: A47253
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-1308 <PLO>
A:Cross-references: GB:L07869; NID:g337359; PIDN:AAB59446.1; PID:g337360
A>Note: sequence extracted from NCBI backbone (NCBIP:126842)
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; growth factor receptor
F:716-981/Domain: protein kinase homology <KIN>
F:724-732/Region: protein kinase ATP-binding motif

Db 541 ECVKQCNIHQGEPRFERDSKLCCHSECLVQNSTAYNTTSCSPGPDHCKMAHFDGPH 600
QY 596 CVARCPGKVPDLSPYMPITWKFDEBACOPCPINCHSCVDLDDKGCPCAEQASPLTSIV 655
Db 601 CVKACPAGVGENDL-VKADANAVCOLCHPNCTRGCKGPLEGCP---NGSKTPSTA 656
QY 656 SAVV-GILVVVGLVGVGILIKRQOKIRKYTMRELLQETELVEPLTPSGAMPNQOMRI 714
Db 657 AGVVGILLCLVVVGLGIGLYLRR-HIVKRTLRLLQERLVEPLTPSGEAPNQOHLRI 715
QY 715 LKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLENTSPKANKELDAYVM 774
Db 716 LKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLENTSPKANKELDAYVM 775
QY 775 AGVGPVYSRLIGLICLTSTVQLVLTQMPYGCILLDHVRENRLGSLQDNLNWCQIAKMS 834
Db 776 ASVDNPHVCRLLIGLICLTSTVQLITQMPYGCILLDHVRENRLGSLQDNLNWCQIAKMS 835
QY 835 YLEDVRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESI 894
Db 836 YLEERRLVHRDLAARNLVKTPQHVKITDFGLAKLLGADKEHYHAEGGKVPKIMMALESI 895
QY 895 LRRRTHQSDVWSYGVTVWELMTFGAKPYDGIIPARIPDLLEKGERLPPOPICTIDVMI 954
Db 896 LHRITHTQSDVWSYGVTVWELMTFGSKPYDGIIPASEISSVLEKGERLPPOPICTIDVMI 955
QY 955 MYCKNMIDSECPRELFVSEFSEMRAROPRVVITQ-NEDLGPASPLDSTFVRSLLDDDD 1013
Db 956 MYCKNMIDSECPRELFVSEFSEMRAROPRVVITQ-NEDLGPASPLDSTFVRSLLDDDD 1015
QY 1014 MGDVLDAEYLVPQOQFCPPDPAPGAGGMVHRRSSSTRSGGDLTLGLPSEEPAPRS 1073
Db 1016 MEDIVDAEYLVPQOQFCPPDPAPGAGGMVHRRSSSTRSGGDLTLGLPSEEPAPRS 1073
QY 1074 PL-----APSEGAGSDVFDGLGMAAGKGLQSLPHDPSPLQRYSEDPTVPLPSET--DG 1126
Db 1042 PLLSLISATSNNSATNCID-----RNGQGHVREDSFVQRYSSDPTGNFLESIDDG 1093
QY 1127 VVAPLTCSPQRPVNVQDVRPOPSPRPGPLPAARPAGATLERAKTSLSPKGVKVDVF- 1185
Db 1094 FL-----PAPVYVQ--LMPKKPS-----TAMVQVQVYNNISL 1124
QY 1186 -----AFGGAIVENPEYLTPOGGAAPPHPPPPAFSPAFDNLVYDQ----- 1225
Db 1125 TAISKLPMDERVQNSHSTAVDNPEYL-----NTQSPKAKTVFESSPYIQSGNHQI 1176
QY 1226 --DPPE-----RGAPPSTFKGTPTAENPEYGLDVP 1254
Db 1177 NLDNDPYQOQFLPNETKPNGLLKVPAAENPEYLRVAAP 1214

Query Match	42.8%	Score 2919.5	DB 2	Length 1308
Best Local Similarity	44.8%	Pred. No. 2.7e-112		
Matches 603	Conservative 185	Mismatches 385	Indels 173	Gaps 28
Qy	9	WGILLALLPGAA---STQVCTGTDMMKRLRPASPETHLMDMLRHLYOGCQVQGNLELY	64	
Db	8	WVWSLLVAGTVQPSQSVCAGTENKLSLSLEQQYRALRKYENCEVVMGNLEITS	67	
Qy	65	LPTNASLSFLQDIOEQVQVYLLIAHNVQRPQLRLVRGTQLPEDNYALAVLDNGDPLN	124	
Db	68	IEHNRDLSFLRSVREVYGVVLVALNQFRLPLENLRIRGTKLVEDYALAIFLYRKDG	127	
Qy	125	NTTPVTGASPGGLRELORLSLTEILKGGVLIQRPQLCYOITILWKDIFHKNQOLAULI	184	
Db	128	NF-----GLOELGKLNLTETILNGVYVDQNKFLCYADTIHWQDIVRNPMPSNLTV	178	
Qy	185	DTNRSRACHPCSPMCKSGRSGESSEDCOSLTRTVCAAGC-ARCKGPLTDCHEQCAAG	243	
Db	179	STNGSSCGGRCHKSCGTG-RWGPPTENHCOTLTRTVCAEQDGRGCGYVSDCCBRECAAG	237	
Qy	244	CTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPVNY	303	
Db	238	CSGPKDTCFACMNFDSGACVTCQPTFVYNPTTFQLEHFNHNAKYTYGAFCVKCKPHNF	297	
Qy	304	LSTDVSGSTLVCPLHNOEVTAEQDTCRCKSPCARVCYGLGMOYIKANSKFTGITELE	363	
Db	298	V-VDSSSVCRACPPSKMEV-EENGIMCKPKCTDICPRACDGTGIGTSLMSAQTVDSNIDK	355	
Qy	364	FAGCKIKFSLAFLPESFDGDPASNTAPQYIKANSKFIGITELTGILYLIWAPDSDLPLS	423	
Db	356	FINCTKINGNLFIWGTGIHGDPYNAIBADPEKLVNVRTVREITGFLNIOSWPPNMTDFS	415	
Qy	424	VFQNLQVIRGRILHNGAYSITLQGLGTSWLGRLSRELGLSGLALIHENTHLCFVHTVPWD	483	
Db	416	VFSNLVTIGRVLVSGLSLLILKQOGLTSQFQSLKISAGNIYITDNSNLCTYHTTNWT	475	
Qy	484	QLFRNPQOALLHTANRDEBCVGEGLACHOLCARGHCWGPPTQCVNCOSFPLRQECVEE	543	
Db	476	TLFSTINQIRVIRDNRKAENCTABGVMCNHLCSSDGCWGPDPQCLSCRRFSRGRICIES	535	
Qy	544	CRVLQGLPREVYNARHCLPCHPEQCP-QNGSVTCFGEADOCVACAHYKDPFPFCVARCPS	602	
Db	536	CNLYDGSFRFBFENGSI CVBQDPQEKMEDEGLLTCHGPGPNCTKCSHFQDGNPCVEKCPD	595	
Qy	603	GVKPDLSYMPIWKPFPDEEGACQPCPINCTHSCVDLDDKGC-----PABQRASPL	651	
Db	596	GLOGANSF--IFKYADDPRECHPCHPNTQGCNGPTSHDCIYYPTWGHSTILPQHAR--TPL	652	
Qy	652	TSIVSAVV-GILLVVLGVWFGILIKRRQOKIRKYTMRLLOETELVEPLTPSGAMPNQ	710	
Db	653	--IAAGVIGLFLIVIVGLTFVAVVVRKSIK-KKRALRRFL-ETELVEPLTPSGTAPNQ	708	
Qy	711	QMRILKETELRKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE	770	
Db	709	QLARILKETELRKVKVLGSGAGFTYKGIWVPEGETVKIPVAIKTLNETTGPKANVFEMDE	768	
Qy	771	AYVMAGVSPVSRLLIGLCTSTVOLVTOLMPYGCLLDHYRENKRLGSDLLNMCWQIA	830	
Db	769	ALIMASMDHPLVRLLGVCLSPTQLVTQLMPHGCLLEYVHEHKDNTGSQULLNMCWQIA	828	
Qy	831	KGMSYLEDLVLRDLAARNVLKSPNHVKITDFGLARLIDIDETEHADGGKVPKIKMA	890	
Db	829	KGMNYLEERLVRDLAARNVLKSPNHVKITDFGLARLLEGEDEKEYNADGGKMPIKMA	888	
Qy	891	LESILRRPFTQSDWNSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTID	950	
Db	889	LECITHYKFTQSDWNSYGVTVWELMTFGGKPYDGIIPTRBIPDLLEKGERLPQPPICTID	948	
Qy	951	YVMIMVKCWMIDSCRRPREFRELVSERMDAPORFVVIQNED-LGPASPLDSTFYRSL	1009	
Db	949	YVMIMVKCWMIDADSRPKFELAAEFERMDAPORYLIVIQGDDRMKLPSPNDSKFFQNLL	1008	
Qy	1010	EDDDMGDLVDAEEYLVPPQGGFCFPDPAPGAGGMVHRRHSRSTSGGGDTLTGLPEPSEE	1069	

RESULT 8

RESULT
S061A2

Subtotal
proteins

proel.
N. 27 + 6

2003-03-25

C;Spec

C; Date

C; Acce

R; Witt.

Nature

A;Titl

A; Refe

A;Acce

A; Mole

A;Resi.

A; CroB

R; Adam

Oncogenes:

A;Titl

A; Refe.

A; Acce

A: Stat

A:Mole

A: Regi

A: CROB

C: Gene

A. G. G.

A: Gene
A: Man

A;Map

A; Inly

Supers:

C; Keyw

F;1-25

Quer-
Best
Matc

Db 123 YQK-NPSSP--DVYQVLKQLQLSNLTELISGVKVSHPNLLCNVETINWWDIVDKTSNP 179
QY 180 ALTIDNRSRACHPCSPMKGRWCWGESSEDCQSLTRTVACAGC-ARCKGPLEPTDCCHE 238
Db 180 TNNLIHPAFERQCKQCDHGCVNGSWAPGPGHGHCKFTKLLCABQCNRRCKGPKPIDCCNE 239
QY 239 QCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTA 298
Db 240 HCAGGCTGPRATDCLACRDNFDGTDCTCPPKIYDIVSHQVVDNPNTKYTFGACVKE 299
QY 299 CPYNYLSTDVGSCTLVCPHLNQVTAEDGTQRCCKSPKPCARVCYGLGMOYIKANSKFTG 358
Db 300 CPYNYVUTE-GACVRSAGMLEVD-ENGKRSCKPCDGVCPKVCIDGIGTSL-SNTIAVN 356
QY 359 ITL-EPAGCKITFGLAFIPESFDGDP---ASNTAPQYIKANSKFTIGITELGYLYISA 414
Db 357 STNIRFSNCKINGDIILNRNSFEGDPHYKIGTMDPEHL---WNLTTVKEITGYLIWM 413
QY 415 WPSLDPDLSPONLQVIRGILHNGAYS-LTLOGLISWLGRLSRLBELSGLALIHNTH 473
Db 414 WPNMTSLVSFQNLRIIRGRTTFSRGFSFVVQVVRHLQWLGLSLKEVSGAGNVILKNTLQ 473
QY 474 LCFVHTVPMDLPFRNPHQALLHTANRPEDECVGEGLAGHQLCARGHCWGPPTQCVNCSQ 533
Db 474 LRYANTINWRLEFRSDQIEYDART-----ENQTCNNECSEDGCV-PGPTMVCVSLH 525
QY 534 FLRGQCEBCEVRLOGLPREYVNAHCLPCHPECPQNGSVTCFGPEADQCACAHKDP 593
Db 526 VDRGRCVASCNLLQEPREAOVDGRVCVQHQECLVQDLSLTCYGPANCSSAHPQDG 585
QY 594 PFCVACPGGVKPDLSYMIWPKPDEEGACOPCINCTHSCVDLDDKGCAPAEORASPLTS 653
Db 586 PQCIPCPHGILGDGDTL-IWKYADRWGQCPCHQCTQCSCPGSLSGCRGD-IVSHSL 643
QY 654 IYSAVVGILLVVLGVVFGILLKRRQKIRKYTMRLLOETELVEPLTSGAMPNQAQR 713
Db 644 AVGLVSLGTLTIVALLIVLLVLRRIK-RKRTIRCLLQELKEVLEPLTSGQAPNAFLR 702
QY 714 ILKETLRKVKVLGSGAFGVYKGIWIPDENVKIPVAIKVLRNTPSKANKELDEAYV 773
Db 703 ILKETEFKDRVLGSGAFGVYKGLWNPDENIRIPVAIKVLEATSPKVQNEVLDEAYV 762
QY 774 MAGVSPYVSRLLIGICLTSTVQLVTQIMPYGCLLDHVRENRLGSGQDLNWCWQIAKM 833
Db 763 MASVDHPHVCRLGICLTSAVQLVTQIMPYGCLLDYVROHERICQGWLLNWCWQIAKM 822
QY 834 SYLEDVRLVHRDLAARNLVKSPNHVKITDPGLARLLIDETEHADGKVPKKNWALS 893
Db 823 NYLEERHLVHRDLAARNVLLKNPNHVKITDFGLSKULLTADKEKYQADGKVPKKNWALS 882
QY 894 ILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICTIDVYM 953
Db 883 ILQWYTHQSDVMSYGVTVWELMTFGSKPYDGIPIAKEIASVLENGERLPQPPICTIEVM 942
QY 954 IMVKWMDISECPRELVSFSPRMARDPQRPVVTQNEBGLPASPLDSTFYVRSLLDD 1013
Db 943 IILKWMIDPSSRPRELVSFSPRMARDPSRYLVIQG---NLPSLSDRRLFSRLSSD 999
QY 1014 MGLVDABEYVLPQOGFFCPDPAPAGGVMVHRHSSRSTSGGDDTLTGLESEEAAPS 1073
Db 1000 --DVVDADYLLPYKRI-----NRQS-----B 1020
QY 1074 PLAPSEGAGSDVPDGLGMAAKGLQSLPHTDPSPLQRYSEDPV-PLPSETDGVVAPLT 1132
Db 1021 PCIPPTGH-----PVRENSITRLNISDPTQNALEKDLGH----- 1055
QY 1133 CSPQRYVNPQDVRPQ-----PSPRE-----GFLP-AARPAGATLERAKTLSPGKG 1179
Db 1056 -----EYVNPQSETSRSLSDIYNPNYEDLTDGWPVSLSSQEAETNFSRPEYINTNQS 1110
QY 1180 VVKDVFAGCAVENPEYLTFQGAAPQPHPPAFSPAFNLYWDDPDPERGAAPPSTFKG 1239
Db 1111 L---PLVSSGSMDDPDP---QAG-----YQAAF-----LPQTGALTGNMFP 1145

QY 1240 TPTAENPEYLG 1250
Db 1146 LPAENLEYLG 1156

RESULT 9

A36223

kinase-related transforming protein (erbB3) (BC 2.7.1.-) precursor - human
Species: Homo sapiens (man)

C>Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000

C.Accession: A36223; 159164

R:Kraus, M.H.; Issing, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.

Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989

A:Title: Isolation and characterization of ERBB3, a third member of the ERBB/epidermal

A:Reference number: A36223; MUID:90083234; PMID:2687875

A:Accession: A36223

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1342 <KRA>

A:Cross-references: GB:M29366

R:Plowman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todaro, G.

Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990

A:Title: Molecular cloning and expression of another epidermal growth factor receptor-r

A:Reference number: 159164; MUID:90311312; PMID:2164210

A:Accession: 159164

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: mRNA

A:Residues: 1-559, 'G', 561-957, 'F', 959-1063, 'G', 1065-1342 <RES>

A:Cross-references: GB:M34309; NID:g183990; PIDN:AAA35979.1; PID:g306841

C:Genetics:

A:Gene: GDB:ERBB3; HER3

A:Cross-references: GDB:119880; OMIM:190151

A:Map position: 12q13-12q13

C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homol

C:Keywords: ATP; phosphotransferase

F:707-972/Domain: protein kinase homology <KIN>

F:715-723/Region: protein kinase ATP-binding motif

Query Match 34.7%; Score 2366.5; DB 2; Length 1342;

Best Local Similarity 39.9%; Pred No. 1.le-89;

Matches 524; Conservative 197; Mismatches 458; Indels 135; Gaps 34;

QY 10 GLLALLPPGAA--STQVCTGTMKRLPASPETHLMLRHLYQGVVQVQGNLELYLPT 67

Db 11 GLLFSLARGSEVGNSSQAVCPGTLNGLSVTGDANQYQTLKYLERCEVVMGNLEIVLTGH 70

QY 68 NASLSFLQDIQEVQGVYLIAHNQVQVFLQRLRIVRGTQLFEDNYALAVLDNGLDPLNNT 127

Db 71 NADLSFLQIREVTGYVLVAMNEFSTLPLNLRVVVRGTQVYDGKFAIFVM----LNynt 125

QY 128 PVTGASPGGLRELQLRSLEILLKGVLIQRPOLCVQDTILWKDIFHKNNQLALTIDTN 187

Db 126 ----NSSHALRQRLTQTEILLSGVYTEKNDKLMCHMDITDWRDIVRDRD---AIVVKD 178

QY 188 RSRAHPSPCMKGRWCWGESSEDCQSLTRTVACAGC-ARCKGPLEPTDCHEQCAAGCTG 246

Db 179 NGRSCPPCHEVCKG-RWGPGESEDCQTLTKTICAPQCNGHCFGNPNQCCHDECAGGCSG 237

QY 247 PKHSCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAAPNYLST 306

Db 238 PDDTDCFACRHFNDSGACVPRCPQPLVYNNKLTFLQEPNPHTKYQYGGVGVASCNPHFV-V 296

QY 307 DVGSCCTLVCPHLNQVTAEDGTQRCCKSPKPCARVCYGLGMOYIKANSKF--IGITELB- 363

Db 297 DQTSVCRACPPDNMEVD-KNGLKMCPCGGLCKPAKEGTG-----SGSRFQTVDSNIDG 350

QY 364 FAGCKKIFGSLAFPESEFDGPPASNTAPQYIKANSKFICITELTGYLYISAWPDSLPLDS 423

Db 351 FVNCIKILGNLDFLTGLNGDPWHKI PALDPEKLVNFRVREITGYLNIQSWPPHMHFNS 410

QY 424 VFQNIQVIRGRILHNGAYS-LTLOGLGISWLGRLSRLBELSGLALIHNTHLCFVHTVPW 482

Db 411 VFSNLTTIGRSYLNRFSLIMKLNVTSLGFRSLKEISAGRIYISANRQLCYHSLNW 470
Qy 483 DQLFRPHQALLHTA-NRPEDECVGGLACHQIACARGCHWGPGPTQCVNCSQFLRGQECV 541
Db 471 TKVLRGTEBERLDIKHNRPRDCVAEGKVCDPICSSGGCGWGPQGLSCRNYSRGVCV 530
Qy 542 EECRVQLGLPREYVYNAHCLPCHPEQOPQNGSVTCFQPEADQCVACAHYKDDPPFCVARCP 601
Db 531 THCNFLNGEPREFAHEACFSCHPECPMEGTATCNGSGSDTCAQCAHFRDGHCVSSCP 590
Qy 602 SGVKPDLSPYMPIWKFPDEBAGQCPINCHTHSCVDLDDKGCPEAQRA----SPLTSIVSA 657
Db 591 HGVLG--AKGPIYKYPDVQNECPCHENCCTQCGKPELQDCLGQTLVLIGKTHLTALTY 648
Qy 658 VVGILLVVLGVFGILIKERQKIR-KYTWRLQLQETELVEPLTPSGAMPNQAQWILK 716
Db 649 IAG--LVVIFMUGGTFLYWRGRRIQNKRAMRYLERGESIEPLDPS-EKANKVLARIFK 705
Qy 717 ETELRLKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDYAVMAG 776
Db 706 ETELRLKVLGSGVFGTVHKVMIPEGESIKIPVCIKVIEDKSGROSFOAVTDHMLAIGS 765
Qy 777 VGSFYVSRLLGICLTSTVQLVLTQMPYGCILLDHVRENRLGSLQDILLNWCQIAKMSYL 836
Db 766 LDHAHIVRLGLGCPGSSQLVTVPLGSLLDHVRQHRGALGPQLLLNMGVQIAKGMYYL 825
Qy 837 EDVRLVHRLAARNVLKSPNHVKITDPGLARLLIDETEHADGGKVPILKNWALESLR 896
Db 826 EEHGMVHRNLAARNVLKSPQVQVADFGVADLLPDDKOLLYSEAKTPILKNWALESIHF 885
Qy 897 RRTTHOSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPPOPICTIDVYMIW 956
Db 886 GKTHQSDVMSYGVTVWELMTFGAEYAGRLAEVDPDLLEKGERLAQPOICTIDVYMWV 945
Qy 957 KCMWIDSECRPRELVSEFSRMARDPQRFVITQNEDLGPA----SPLDSTFYHSLLEDD 1013
Db 946 KCMWIDENIRPTEKELANETRMARDPPRYLVIKRES-GPIAGPGEHPHGLTNKLEEEV 1004
Qy 1014 MGLVDAAEEVLVQQGFFCDPAPGAGMVHHRSSSTRSGGDLTLGLEP-SESEAP 1072
Db 1005 LEPELDLDLEAED-----NLATTGLSALSLPVTPLSPGLEDDEVDNG 1158
Qy 1073 SPLASEGAGSDVFDGLGMAKGLQLPHTD-PSPLQRYSEDPTVPLP-----SETD 1125
Db 1046 SLSPSSGY-MPMNQNLGESCQESAVSGSERCPVSLH-----PMRGLGASESE 1098
Qy 1126 GYVA-----PLTCSPOE----YVNPQDVVRPQPPPREG----- 1156
Db 1099 GHVTGEAELOEKVSMCRSRSRSPRPRGDSAYHSQRHSLLTPVTPLSPGLEDDEVDNG 1158
Qy 1157 --LPAARPAGATLERAKTLP-SKNGV-----KDVFAFGGAVENPEYLTPOGGAAPQ 1207
Db 1159 YNMPDTHLKTSPSREGTUSVGLSSVLGTEEBED-----EEYENRRRRHS-P 1209
Qy 1208 HPPAPSPADNLYYWD-----ODPPERGAPPSTFKGTPTAENPEYL 1249
Db 1210 HPPRPSLEELGEYMDVGDLSASLGSTQSCPLHPVIMPATAGTTPDEDEYEM 1263

RESULT 10
JC4387
epidermal growth factor receptor; homolog precursor - rat
N;Alternate names: ErbB3 protein; HER3 protein
C;Species: Rattus norvegicus (Norway rat)
C;Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998
C;Accession: JC4387
R;Hellyer, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.
Gene 165, 279-284, 1995
A;Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protein.
A;Reference number: JC4387; MUID:96096535; PMID:8522190
A;Accession: JC4387
A;Molecule type: mRNA
A;Residues: 1-1339 <HEL>

A;Cross-references: GB:U29339; NID:g915389; PID:g915390
A;Experimental source: liver
A;Note: The authors translated the codon AAC for residue 369 as Thr and GTT for residue
C;Comment: this protein is a functional heregulin receptor that transduces signals to ti
C;Genetics:
A;Gene: ErbB3
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homol
C;Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane protein
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1339/Product: epidermal growth factor homolog #status predicted <TM>
F;640-659/Domain: transmembrane #status predicted <TM>
F;705-970/Domain: protein kinase homology <KIN>
F;713-721/Region: protein kinase ATP-binding motif
F;939, 1051, 1156, 1194, 1196, 1219, 1257, 1259, 1273, 1286, 1325/Binding site: phosphate (Tyr)
Query Match 33.5%; Score 2280.5; DB 2; Length 1339;
Best Local Similarity 40.1%; Pred. No. 3.6e-86;
Matches 515; Conservative 173; Mismatches 438; Indels 159; Gaps 35;
Qy 3 LAALCRWGLLLALLPPGAA---STOVCTGTDMLRLPASPTHLDMLRHLHYQGCQVQGN 59
Db 7 LQVLC----FLSLARGSEMNSQAVCPGTLNGLSVTGADNQYQTLYKLYEKCEVVMGN 62
Qy 60 LELTYLPTNASISFLQDIQEVQGYVLIHQNVRQVPLQRLRIVRGTLFEDNYALAVLDN 119
Db 63 LEIVLTGHNADLSFLQWIREVTAYVLVAMNFSVLPLNLRVVRGTQVYDGKFAIFVM-- 120
Qy 120 GDLNNTPTVTGASPGGLRELQRLSITELTKGVLQIQRNPOLCYQDTILWKDIFHKQNQL 179
Db 121 ---LNYNT---NSSHALRQLKFTQITELSGVVIETKNDKLCMDITDMDRIVR-- 170
Qy 180 ALTLDTNRSRACHPCSPMCKSGWGESSEDCQSLTRTVACAGG-ARCKGLPLPTDCHE 238
Db 171 GAEIVVYKANGANCPCHVECKG-RCWGPDPDCCQLTKTICAPQNGRCFGPNQCCHD 229
Qy 239 QCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFFSMPNPEGRYTFGASCVTA 298
Db 230 ECAGGCSGQDPTDCACRFNDSGACVPRCPPEPLVYNKLTFLQEPNPHTKYQVGGVCVAS 289
Qy 299 CPYNYLSTDVGCTLVCPLNQEVTAEDGTQRCCKSKPCARVCYGL--GMOYIKANSKF 356
Db 290 CPNFFV-VQDTFCVRACPPDKMEVD-KHGLKNCEPGGLCPKACEGTGSGSYQTVDDSN 347
Qy 357 IGITELEFAGCKIKFGLSLAFLPESFDGDPASNTAPQYIKANSKFITELTGLVLYISAWP 416
Db 348 ID----GFVNCYKILGNLDFLITGLNVDPWHKIPALDPEKLVNFRVTRVITGLNIQSWP 403
Qy 417 DSLPDLVSFQNLQVIRGRILHNAYS-LTLOGLGISWLGRLSRLRELGSGLALIHNTILC 475
Db 404 PHMNFVSFNLTITIGRSLYNRGFSLLIMKLNVTSLGFRSLKEISAGRVVISANQQLC 463
Qy 476 FVHTVPMDQLFRPHQALLHTA-NRPEDECVGGLACHQIACARGCHWGPGPTQCVNCSQF 534
Db 464 YHHSNLNWTLLRGPSEERLDIKYDRPLGCELAEGKVCDDPLCSGGCGWGPQGLSCRN 523
Qy 535 LSGQCEVEECRVQLPREYVYNAHCLPCHPEQOPQNGSVTCFQPEADQCVACAHYKDDP 594
Db 524 SREGVCVTHCNFLQGEPRFVHEAQCFCHPCLPMEGTSTYNGSGSDACACAHFRDGP 583
Qy 595 FCVARCPGSKVPDLGYMPIWKFPDEBAGQCPINCHTHSC--VDLDDKGCPEAQRASPLT 652
Db 584 HCVNSCPHGILG--AKGPIYKYPDAQNECRPCHENCCTQCGNGPELQDCLGQAEVLMSKPH 641
Qy 653 STVSAVVGILLVVLGVFGILIKERQKIR-KYTWRLQLQETELVEPLTPSGAMPNQAQ 711
Db 642 LVIAVTVG--LAVILMILGGSFLYWRGRRIQNKRAMRYLERGESIEPLDPS-EKANKVL 698
Qy 712 MELKETELRKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDYEA 771
Db 699 ARIFKETELRKVLKVLGSGVFGTVHKVMIPEGESIKIPVCIKVIEDKSGROSFOAVTDM 758
Qy 772 YVMAGVGSYPVSRLLGICLTSTVQLVLTQMPYGCILLDHVRENRLGSLQDILLNWCQIAK 831

Matches 345; Conservative 69; Mismatches 122; Indels 92; Gaps 15;

```
QY 578 GPADQCVCAHYKDPFCVARGSPGVKPLSYMPIWKFPDEGACQPCPINCINCHSCVDL 637
Db 1 GP--DHCMKCAHFDGPHCVKACPAAGVLGENDTL-VMKYADANAVCOLCHPNCTRGCKGP 57
QY 638 DDGCPAEQASPLTSTVSAAV-GILLVVVLGVVFGILLIKRRQOKIRKYTMRLLOETEL 696
Db 58 GLGEGCP---NGSKTPSIAAGVGGLLCLVVGGLGIGLYLRR-HIVKRTLRLLQEREL 113
QY 697 VEPLTPSGAMPNOAQRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRL 756
Db 114 VEPLTSGEAPNOAHLRIKETEFKVKVLGFGAFGVYKGLIPEGEKVTIPVAIKELR 173
QY 757 ENTSPKANKILDEAYMAGVSGPYVSRLLIGICLTSTVQLVTLMPYGCLLDHVRENRR 816
Db 174 EATSPKANKILDEAYMASVDNPHVCRLLIGICLTSTVQLITQLMPYGCLLDIYREHK 233
QY 817 LGSODLLNCMOJAKGMSYLEVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETE 876
Db 234 IGSQYLLNWCQJAKGMNLYLEERHLVHRDLAARNVLKTPQDVKITDFGLAKQLGADEKE 293
QY 877 YHADGGKVPPIKMALESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLE 936
Db 294 YHAEKGKVPPIKMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIPIASEIS 353
QY 937 KGBRLPQPTCTIDVYIMVKMWIDSECRPRFRELSEFSRWARDPQRFVVIQ-NEDLG 995
Db 354 KGBRLPQPTCTIDVYIMVKMWSDADSRPKFRELIAEFKSMARDPPRYLVIQDERMH 413
QY 996 PASPLDSTFVRSILLEDDMGDLVDAEYLVPOGFCFDPAPGAGGMVHRRSSSTRSG 1055
Db 414 LPSFTDSKFTLMEEDMEDIVDADEYLVPHQGF-----NSPST---454
QY 1056 GGLTLGLEPSEEAEPSP-----APSEGAGSDVFDGDLGMGAAGLQSLPHTDPSPLQ 1110
Db 455 -----SRTPLLSLSATSNNSATNCIDRNG-----H-----481
QY 1111 RYSEDTPTVLPSTGDYVAPLTCSPQPEYVNPQDVPQPPSPREGPLPAARPAGAT-LER 1169
Db 482 -----PVREDGFL-----PAPEYVNO--LMPKPKSTAMVQNIYVYISLTAISK 523
QY 1170 AKTLPKGVGVKDVFAFGAVENPEYL 1197
Db 524 LPMDSRYQ-----SHSTAVDNPEYL 544
```

RESULT 15

B44776

protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain ES4)

C;Species: avian erythroblastosis virus

C;Date: 28-Apr-1993 #sequence_revision 28-Apr-1993 #text_change 04-Feb-2000

C;Accession: B44776

R;Bruskin, A.; Jackson, J.; Bishop, J.M.; McCarley, D.J.; Schatzman, R.C.

Oncogene 5, 15-24, 1990

A;Title: Six amino acids from the retroviral gene gag greatly enhance the transforming F

A;Reference number: A44776; PMID:90206603; PMID:1969616

A;Accession: B44776

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-540 <BRU>

A;Cross-references: GB:X52211

C;Genetics:

A;Gene: erbB

C;Superfamily: epidermal growth factor receptor; protein kinase homology

C;Keywords: ATP; phosphotransferase; tyrosine-specific protein kinase

F;130-395/Domain: protein kinase homology <KIN>

F;138-146/Region: protein kinase ATP-binding motif

Query Match

23.8%; Score 1623; DB 2; Length 540;

Best Local Similarity 54.9%; Pred. No. 1.1e-59;

Matches 340; Conservative 69; Mismatches 120; Indels 90; Gaps 14;

Search completed: July 22, 2003, 09:09:39

Job time : 30.9062 secs

```
QY 587 CAHYKDPFCVARGSPGVKPLSYMPIWKFPDEGACQPCPINCINCHSCVDLDDKGCQPAEQ 646
Db 3 CAHFIDGPHCVKACPAAGVLGENDTL-VMKYADANAVCOLCHPNCTRGCKGPGLGCP---58
QY 647 RASPLTSTVSAVV-GILLVVVLGVVFGILLIKRRQOKIRKYTMRLLOETELVEPLTPSGA 705
Db 59 NSKTPSIAAGVGGLLCLVVGGLGIGLYLRR-HIVKRTLRLLQERELVEPLTPSGE 117
QY 706 MPNOAQRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANK 765
Db 118 AFNOAHLRIKETEFKVKVLGFGAFGVYKGLIPEGEKVTIPVAIKELRENTSPKANK 177
QY 766 EILDAYMAGVSGPYVSRLLIGICLTSTVQLVTLMPYGCLLDHVRENRRGSLQDILLW 825
Db 178 EILDAYMAGVSGPYVSRLLIGICLTSTVQLITQLMPYGCLLDIYREHKDNIQSQYLLNW 237
QY 826 CMQJAKGMSYLEVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEYHADGGKVP 885
Db 238 CVQJAKGMNLYLEERHLVHRDLAARNVLKTPQDVKITDFGLAKQLGADEKEYHADGGKVP 297
QY 886 IKWMALESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLEKGERLPQPP 945
Db 298 IKWMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIPIASEISSVLEKGERLPQPP 357
QY 946 ICTIDVYIMVKMWIDSECRPRFRELSEFSRWARDPQRFVVIQ-NEDLGASPLDSTF 1004
Db 358 ICTIDVYIMVKMWSDADSRPKFRELIAEFKSMARDPPRYLVIQDERMHLPSPDTSKF 417
QY 1005 YESLLEDDMGDLVDAEYLVPOGFCFDPAPGAGGMVHRRSSSTRSGGGLTLGLE 1064
Db 418 YRLMEEDMEDIVDADEYLVPHQGF-----NSPST-----449
QY 1065 PSEEEAPRSFL-----APSEGAGSDVFDGDLGMGAAGLQSLPHTDPSPLQRYSEDTVP 1119
Db 450 -----SRTPLLSLSATSNNSATNCIDRNG-----H-----476
QY 1120 LPSETDGYVAPLTCSPQPEYVNPQDVPQPPSPREGPLPAARPAGAT-LERAKTLPQGN 1178
Db 477 -PVREDGFL-----PAPEYVNO--LMPKPKSTAMVQNIYVYISLTAISKLPMSRYQN 527
QY 1179 GVVKDVFAFGAVENPEYL 1197
Db 528 -----SHSTAVDNPEYL 539
```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:11:38 ; Search time 37.9774 Seconds
(without alignments)
4403.399 Million cell updates/sec

Title: SEQ4-250-264-12
Perfect score: 6801
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYGLDVPV 1255

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6720	98.8	1255	21	AAV92620 Human herégulin 2
2	6720	98.8	1255	22	AAE12130 Human tyrosine kin
3	6720	98.8	1255	22	AAE60167 HER2 transgene pla
4	6720	98.8	1255	23	AAU74545 Human HER2 (Erbb2)
5	6714	98.7	1255	17	AAW01111 HER-2/neu protein.
6	6714	98.7	1255	20	AAW92406 Human HER-2/neu pr
7	6714	98.7	1255	21	AAE21198 Human HER-2/neu pr
8	6714	98.7	1255	21	AAV84780 Amino acid sequenc
9	6714	98.7	1255	21	AAE85458 Human HER-2/neu pr
10	6714	98.7	1255	22	AAE88267 HER2/neu amino aci

11	6714	98.7	1255	23	AAE24067 Human Her-2 protei
12	6714	98.7	1255	23	AAE20479 Human Her-2/neu pr
13	6714	98.7	1255	23	AAE51143 Human Her-2/neu on
14	6714	98.7	1255	23	AAU77114 Human Her-2/neu po
15	6671	98.1	1433	14	AAE39568 Sequence of c-erbB
16	6550	96.3	1223	23	AAU98923 Human breast cance
17	6397	94.1	1200	21	AAE21208 Human HER-2/neu pr
18	5915.5	87.0	1256	21	AAE21199 Rat HER-2/neu prot
19	5915.5	87.0	1256	23	AAE51144 Mouse Her-2/neu on
20	5889.5	86.6	1256	21	AAE21206 Amino acid sequenc
21	5889.5	86.6	1256	22	AAE62860 Mouse Her-2/neu pr
22	5889.5	86.6	1256	23	AAE51151 Mouse Her-2/neu on
23	4800	70.6	919	21	AAE21203 Human HER-2/neu fu
24	4800	70.6	919	23	AAE51148 Her-2/neu extracel
25	4039.5	59.4	920	23	AAE51152 Mouse Her-2/neu ex
26	4039.5	59.4	926	23	AAE51153 Mouse Her-2/neu ex
27	3684	54.2	712	21	AAE21204 Human HER-2/neu fu
28	3684	54.2	712	23	AAE51149 Her-2/neu extracel
29	3538	52.0	782	18	AAW19764 Her2-GM-CSF immuno
30	3536	52.0	653	21	AAE21200 Extracellular HER-
31	3536	52.0	653	23	AAE51145 Human Her-2/neu on
32	3498	51.4	645	22	AAE60408 Human ErbB2 oncopr
33	3498	51.4	645	22	AAE61593 Human ErbB2 extrac
34	3433	50.5	951	21	AAE44993 DC8ecFv-erbB2EC fu
35	3330	49.0	624	11	AAE08222 Extracellular port
36	3121	45.9	1210	21	AAE19259 Amino acid sequenc
37	3121	45.9	1210	21	AAE19259 Human EGF receptor
38	3121	45.9	1210	23	AAE23019 Human Her-1 protei
39	3121	45.9	1210	23	AAE50768 Human epidermal gr
40	3119	45.9	1210	22	AAE68420 Amino acid sequenc
41	3084	45.3	583	23	AAE20483 Human protein for
42	3084	45.3	587	23	AAE20481 Human protein for
43	3083	45.3	589	23	AAE20484 Human protein for
44	3083	45.3	600	23	AAE20482 Human protein for
45	3080	45.3	1210	23	AAE51768 Human epidermal gr

ALIGNMENTS

RESULT 1
AAV92620
ID AAV92620 standard; Protein; 1255 AA.
XX
AC AAV92620;
DT 10-AUG-2000 (first entry)
DT
XX Human herégulin 2 (Her2).
DE
XX Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;
KW self-protein; cancer; breast cancer; prostate cancer;
KW cell-associated peptide antigen; foreign epitope.
XX Homo sapiens.
OS
XX Key Location/Qualifiers
FH Domain /label= N-terminal
FT /note= "mature polypeptide"
FT 5..25
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 59..73
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 103..117
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 149..163
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 174..323
FT Domain

FT Region /label= Cysteine_rich_domain
 FT 210..224
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 250..264
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 324..483
 FT /label= Ligand_binding_domain
 FT 325..339
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 369..383
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 465..479
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 484..623
 FT /label= Cysteine_rich_domain
 FT 579..593
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 624..654
 FT /label= Transmembrane_domain
 FT 632..652
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 653..667
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 655..1010
 FT /label= Tyrosine_kinase_domain
 FT 661..675
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 695..709
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 710..730
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT 1011..1235
 FT /label= C-terminal_domain
 XX WO200020027-A2.
 XX 13-APR-2000. 99WO-DK00525.
 XX 05-OCT-1999;
 XX 05-OCT-1998; 98DK-0001261.
 XX 20-OCT-1998; 98US-0105011.
 XX (MEBI-) M & E BIOTECH AS.
 XX Steinaa L, Mouritseen S, Nielsen KG, Haaning J, Leach D, Dalum I;
 XX Gautam A, Birk P, Karlsson G;
 XX WPI; 2000-349917/30.
 XX N-PSDB; AAA09455.
 XX Inducing immune responses to weakly immunogenic, tumor associated
 XX peptide antigens for the treatment of breast and prostate cancer
 XX
 XX Claim 62; Page 193-198; 220pp; English.
 XX This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of
 XX Her2 can be used in the claimed method as an autovaccine to induce a CTL
 XX response. Subdominant CTL epitopes, antibody binding regions and
 XX cysteine residues involved in disulfide bonds are preserved in the
 XX immunogenized forms. Regions suitable for the insertion of foreign T
 XX helper epitopes were identified (see features table). The method

CC is used for inducing immune responses against weakly immunogenic
 CC cell-associated peptide antigens (PA) such as those associated with
 CC cancers (self-proteins), e.g. human prostate specific membrane antigen
 CC (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).
 CC The method comprises effecting simultaneous presentation by antigen
 CC producing cells (APCs) of the animals immune system of: (1) at least 1
 CC CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1
 CC B-cell group derived from the cell-associated PA; and (2) at least 1
 CC first T-helper cell group which is foreign to the animal. Analogues of
 CC human PSM, human Her2 and human/murine FGF8b comprising a substantial
 CC part of all known and predicted CTL and B-cell epitopes of the respective
 CC PA and including at least one foreign T helper epitope are also claimed.
 CC The method is used to treat prostate, prostate/breast or breast cancer
 CC when the PA is human PSM, FGF8b and Her2, respectively.
 XX
 XX SQ Sequence 1255 AA;

Query Match 98.8%; Score 6720; DB 21; Length 1255;
 Best Local Similarity 99.0%; Pred. No. 0;
 Matches 1242; Conservative 2; Mismatches 11; Indels 0; Gaps 0;
 QY 1 MELAALCRWGLLLALLPPGAASSTQVCTGDMKRLPASPETHLDMRLHLYQGCVVQGNL 60
 DB 1 MELAALCRWGLLLALLPPGAASSTQVCTGDMKRLPASPETHLDMRLHLYQGCVVQGNL 60
 QY 61 ELTYLPTNASLSFLQDIQEVQGVLIHQNQVROVPLQRLRIVRGTLQFEDNYALAVLDNG 120
 DB 61 ELTYLPTNASLSFLQDIQEVQGVLIHQNQVROVPLQRLRIVRGTLQFEDNYALAVLDNG 120
 QY 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILMKDIFHKNNQLA 180
 DB 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILMKDIFHKNNQLA 180
 QY 181 LTLIDNRSRACHPCSPMCKGSRGWESSEDCOSLRTTVCAGGCARCKGSLPTDCCHQEC 240
 DB 181 LTLIDNRSRACHPCSPMCKGSRGWESSEDCOSLRTTVCAGGCARCKGSLPTDCCHQEC 240
 QY 241 AGCTGPKHSDCLACLFHNSHGICELHCPALVOYIKANSKFIGITELRYTFGASCVTACP 300
 DB 241 AGCTGPKHSDCLACLFHNSHGICELHCPALVTYNTDTFESMPNPGRTYTFGASCVTACP 300
 QY 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQCEKCKPCARVCYGLGMEHLREVRVTSAN 360
 DB 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQCEKCKPCARVCYGLGMEHLREVRVTSAN 360
 QY 361 IQEFAGCKKI FGS LAF L P E S F D G D P A S N T A P L Q E Q L V F E T L E E I T G V L Y I S A W P D S L P 420
 DB 361 IQEFAGCKKI FGS LAF L P E S F D G D P A S N T A P L Q E Q L V F E T L E E I T G V L Y I S A W P D S L P 420
 QY 421 DLSVFQNLQVIRGRILHNGAYS L T L Q L G I S W L G L R S L R E L G S G L A L I H N H T L C F V H T V 480
 DB 421 DLSVFQNLQVIRGRILHNGAYS L T L Q L G I S W L G L R S L R E L G S G L A L I H N H T L C F V H T V 480
 QY 481 PWDQLFRNP HQALLHTANRPEDECVGEG LACHQ L C A R G H C W G P G T Q C V N C S Q F L R G Q E C 540
 DB 481 PWDQLFRNP HQALLHTANRPEDECVGEG LACHQ L C A R G H C W G P G T Q C V N C S Q F L R G Q E C 540
 QY 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADOCVCAHYKDPFCVARC 600
 DB 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADOCVCAHYKDPFCVARC 600
 QY 601 PSQVKPDL S Y M P I W K F P D E E G A C O P C P I N C T H S C V D L D D K G C P A E R A S P L T S I V S A V V G 660
 DB 601 PSQVKPDL S Y M P I W K F P D E E G A C O P C P I N C T H S C V D L D D K G C P A E R A S P L T S I V S A V V G 660
 QY 661 ILLVVVLGVVFGILIKRRQKKIRKYTMRELLQETELVEPLTPSGAMPNQAQMRILKETEL 720
 DB 661 ILLVVVLGVVFGILIKRRQKKIRKYTMRELLQETELVEPLTPSGAMPNQAQMRILKETEL 720
 QY 721 RKVKVLGSAFGTVYKGIWIPGENVKIPVAIKVLRNTSPKANKELID EAYMAGV GSP 780
 DB 721 RKVKVLGSAFGTVYKGIWIPGENVKIPVAIKVLRNTSPKANKELID EAYMAGV GSP 780

QY 781 YVSRLLGICLTSTVQLTQMPYGCCLLDHYENRGRGLSGDILLNWCQIAKGSYLEVDV 840
 Db 781 YVSRLLGICLTSTVQLTQMPYGCCLLDHYENRGRGLSGDILLNWCQIAKGSYLEVDV 840
 QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGGKVPKIMWALESILRRFT 900
 Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGGKVPKIMWALESILRRFT 900
 QY 901 HQSDVMSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPICTIDVYIMVKWM 960
 Db 901 HQSDVMSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPICTIDVYIMVKWM 960
 QY 961 IDSECRPRFELVSESRMARDQRFVVIQNEIDLGASPLDSTFYRSLLEDDMDGLVDA 1020
 Db 961 IDSECRPRFELVSESRMARDQRFVVIQNEIDLGASPLDSTFYRSLLEDDMDGLVDA 1020
 QY 1021 EYLVPOQGFCDPAPGAGWVHRHRSSTSRGGDLTLGLEPSEERAPRPLAPSEG 1080
 Db 1021 EYLVPOQGFCDPAPGAGWVHRHRSSTSRGGDLTLGLEPSEERAPRPLAPSEG 1080
 QY 1081 AGSDVFDGDLGMAAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPOPEYV 1140
 Db 1081 AGSDVFDGDLGMAAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPOPEYV 1140
 QY 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPKNGVVKDVFAPGGAVENPEYLTQ 1200
 Db 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPKNGVVKDVFAPGGAVENPEYLTQ 1200
 QY 1201 GGAAPQHPPPAFSPAFDNLVYWDQPPPERGAPPSTFKGPTAENPEYLGLDVVPV 1255
 Db 1201 GGAAPQHPPPAFSPAFDNLVYWDQPPPERGAPPSTFKGPTAENPEYLGLDVVPV 1255

RESULT 2

AAE12130
 ID AAE12130 standard; Protein; 1255 AA.
 XX AAE12130;
 AC AAE12130;
 XX AAE12130;
 DT 18-DEC-2001 (first entry)
 XX Human tyrosine kinase-type receptor, HER-2.
 DE
 XX Therapeutic compound; major histocompatibility complex; vaccine;
 KW antigenic peptide; MHC; immunoregulatory; immune response; HER-2;
 KW adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;
 KW antigen presenting cell; human; tyrosine kinase-type receptor.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH Region 774..782
 FT /note= "Antigenic epitope"
 FT
 XX WO200168677-A2
 XX
 XX 20-SEP-2001.
 XX
 XX 16-MAR-2001; 2001WO-US40328.
 XX
 XX 16-MAR-2000; 2000US-0527487.
 PR
 XX (GENZ) GENZYME CORP.
 XX
 XX Nicolette CA;
 XX
 XX WPI; 2001-616284/71.
 DR N-PSDB; AAD19731.
 XX
 XX Novel synthetic therapeutic compound for inducing immune response and
 PT for use in adoptive immunotherapy, has enhanced binding to major
 PT histocompatibility molecules and enhanced immunoregulatory properties
 PT

XX

Claim 4; Page 63-67; 69pp; English.

XX The invention relates to synthetic therapeutic compounds (antigenic
 XX peptides) with enhanced binding to major histocompatibility complex
 CC (MHC) molecules and enhanced immunoregulatory properties relative
 CC to their natural counterparts. Compounds of the invention are useful
 CC for inducing an immune response in a subject and for use in adoptive
 CC immunotherapy. They are useful as components of anti-cancer vaccines
 CC and to expand immune effector cells that are specific for cancers
 CC characterised by expression of the breast cancer antigen, HER-2.
 CC Polynucleotides that encode peptides of the invention are useful as
 CC hybridisation probes and as primers for the detection of genes of
 CC transcripts that are expressed in antigen presenting cells (APCs), to
 CC confirm transduction of polynucleotides into host cells. The present
 CC sequence is human tyrosine kinase-type receptor, HER-2. Compounds
 CC of the invention are designed based on the HER-2 antigenic peptide
 CC (774-782).

SQ Sequence 1255 AA;

Query Match 98.8%; Score 6720; DB 22; Length 1255;
 Best Local Similarity 99.0%; Pred. No. 0;
 Matches 1242; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLLALLPPGAASVTCTGTDMLRLPASPETHLDMLRLHYQGCVVQGNL 60
 Db 1 MELAALCRWGLLLALLPPGAASVTCTGTDMLRLPASPETHLDMLRLHYQGCVVQGNL 60
 QY 61 ELTYLPTNASLSFLQDIQEVQGYVLIHNVQVQLRLIRVGTQLFEDNVALAVLDNG 120
 Db 61 ELTYLPTNASLSFLQDIQEVQGYVLIHNVQVQLRLIRVGTQLFEDNVALAVLDNG 120
 QY 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQNPQLCVQDTILWKDIFHKNNOLA 180
 Db 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQNPQLCVQDTILWKDIFHKNNOLA 180
 QY 181 LTLIDNRSFACHPCSPCKSGKSGESSEDQSLTRTVACGGCARCKGPLETDCCHEQC 240
 Db 181 LTLIDNRSFACHPCSPCKSGKSGESSEDQSLTRTVACGGCARCKGPLETDCCHEQC 240
 QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVQYIKANSKFGITELRITFGASCVTACP 300
 Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVQYIKANSKFGITELRITFGASCVTACP 300
 QY 301 YNYLSTDVGSCTLVCPLNHNEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRAVTSAN 360
 Db 301 YNYLSTDVGSCTLVCPLNHNEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRAVTSAN 360
 QY 361 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEETGYLYISAWPDSLP 420
 Db 361 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEETGYLYISAWPDSLP 420
 QY 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWGLRSLRELGSGLALIHNNTHLCFVHTV 480
 Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWGLRSLRELGSGLALIHNNTHLCFVHTV 480
 QY 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHOLCARGHCWGPGPTQCVCNCSOFLRGQSC 540
 Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHOLCARGHCWGPGPTQCVCNCSOFLRGQSC 540
 QY 541 VEECRVLOGLPREYVYNARHCLPCHPECOFQNGSVTCFGEADQCACAHYKDPFPFCVARC 600
 Db 541 VEECRVLOGLPREYVYNARHCLPCHPECOFQNGSVTCFGEADQCACAHYKDPFPFCVARC 600
 QY 601 PSGVKPDLISYMPIWKFPDEEGACQPCPNCTHSCVDLDDKGPAPORASPLTSIVSAVVG 660
 Db 601 PSGVKPDLISYMPIWKFPDEEGACQPCPNCTHSCVDLDDKGPAPORASPLTSIVSAVVG 660
 QY 661 ILLVVVLGVVFGILIKRROOKIRKYMRLLOETELVEPLTPSGAMPNQAOQRILKETEL 720
 Db 661 ILLVVVLGVVFGILIKRROOKIRKYMRLLOETELVEPLTPSGAMPNQAOQRILKETEL 720

721 RKVKVLSGAGFTYKGIWIPDGENVKIPVAIKVLRNTPSKANKELDEAYVMAGVSP 780
721 RKVKVLSGAGFTYKGIWIPDGENVKIPVAIKVLRNTPSKANKELDEAYVMAGVSP 780
781 YVSRLLGICLTSTVQLVTLQMPYGLLDHVRNRRGRSGDQLLWNCQIAKMSYLEDVR 840
781 YVSRLLGICLTSTVQLVTLQMPYGLLDHVRNRRGRSGDQLLWNCQIAKMSYLEDVR 840
841 LVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRFT 900
841 LVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRFT 900
901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMVKWM 960
901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMVKWM 960
961 IDSECRPRFRELVSERFARMARDQRFVVIQNEIDLGASPLDSTFYRSLLEDMDGLVDA 1020
961 IDSECRPRFRELVSERFARMARDQRFVVIQNEIDLGASPLDSTFYRSLLEDMDGLVDA 1020
1021 EYLVPOQGFPCPDPAAGAGVHRRHRSSTRSGGGDLTLGLEPSEERAPRPLAPSEG 1080
1021 EYLVPOQGFPCPDPAAGAGVHRRHRSSTRSGGGDLTLGLEPSEERAPRPLAPSEG 1080
1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAPGGAVENPEYLTPO 1200
1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAPGGAVENPEYLTPO 1200
1201 GGAPOHPHPPAPSPAFDNLVYWDQDPPERGAPPSTFKGPTAENPEYLGLDVVP 1255
1201 GGAPOHPHPPAPSPAFDNLVYWDQDPPERGAPPSTFKGPTAENPEYLGLDVVP 1255

RESULT 3

AAB60167
ID AAB60167 standard; Protein; 1255 AA.

XX AC AAB60167;

XX DT 03-APR-2001 (first entry)

XX DE HER2 transgene plasmid construct encoded protein.

XX KW Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
XX KW antibody.

XX OS Homo sapiens.

XX OS Synthetic.

XX PN WO200100244-A2

XX PD 04-JAN-2001

XX PF 23-JUN-2000; 2000WO-US17229.

XX PR 25-JUN-1999; 99US-0141316.

XX PR 16-MAR-2000; 2000US-0189844.

XX PA (GETH) GENENTECH INC.

XX PI Erickson S, Schwall R;

XX DR WPI; 2001-061962/07.

XX DR N-PSDB; AAF24297.

XX PT Treating tumors, particularly breast cancers, which overexpress an ErbB
XX PT receptor and does not respond to an anti-ErbB antibody, comprises
XX PT conjugating the antibody to a maytansinoid -

PS Example 3; Fig 4; 92pp; English.

XX CC The present invention provides a method of treating cancer by
XX CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In
XX CC particular, the antibody is directed against ErbB2 (also known as HER2
XX CC and p185neu). The method is particularly useful in the treatment of
XX CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
XX CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.

XX SQ Sequence 1255 AA;

Query Match 98.8%; Score 6720; DB 22; Length 1255;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 1242; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLLALLPPGAASCTGCTDMKRLRLPASPETHLDMLRHLYQGCVVQGNL 60
DB 1 MELAALCRWGLLLALLPPGAASCTGCTDMKRLRLPASPETHLDMLRHLYQGCVVQGNL 60
QY 61 ELTYLPTNASLSFLQDIQEVQGVLIHNRQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSFLQDIQEVQGVLIHNRQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
QY 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQRPQLCYQDTILWKDIFHKNNQLA 180
DB 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQRPQLCYQDTILWKDIFHKNNQLA 180
QY 181 LTLIDNRRACHPSPCKGRCWGESSEDCOSLTRTVGAGCARCKGPLPDDCCEQC 240
DB 181 LTLIDNRRACHPSPCKGRCWGESSEDCOSLTRTVGAGCARCKGPLPDDCCEQC 240
QY 241 AAGCTPKHSDCLACLFHNSHGICELHCPALVQYIKANSKFITELRVTFFGASCVTACP 300
DB 241 AAGCTPKHSDCLACLFHNSHGICELHCPALVQYIKANSKFITELRVTFFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPHLNHOEVTAEQGTQCEKSKPCARVCYGLGMEHLREVRVTSAN 360
DB 301 YNYLSTDVGSCTLVCPHLNHOEVTAEQGTQCEKSKPCARVCYGLGMEHLREVRVTSAN 360
QY 361 IQEFAGCKKI FGSFLAPLPESPDGDPASNTAPLOPEQLQVFETLEEITGVLYISAWPDSL 420
DB 361 IQEFAGCKKI FGSFLAPLPESPDGDPASNTAPLOPEQLQVFETLEEITGVLYISAWPDSL 420
QY 421 DLSVFONLQVIRGRIHNGAYSITLQGLIGISWGLRSRLRELGLSLALIHNNHLCFVHTV 480
DB 421 DLSVFONLQVIRGRIHNGAYSITLQGLIGISWGLRSRLRELGLSLALIHNNHLCFVHTV 480
QY 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQLCARGHCWGPGTQCVCNCSQFLRGQEC 540
DB 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQLCARGHCWGPGTQCVCNCSQFLRGQEC 540
QY 541 VEECRVLQGLPREYVNHARCLPCHPECPQNGSVTCFGEADQCVAHAKDPFPCVARC 600
DB 541 VEECRVLQGLPREYVNHARCLPCHPECPQNGSVTCFGEADQCVAHAKDPFPCVARC 600
QY 601 PSGVKPDLSPMPTWKPDEEGACOPCPINCTHSCVDLDDKGCPEARSPALTSIVSAVVG 660
DB 601 PSGVKPDLSPMPTWKPDEEGACOPCPINCTHSCVDLDDKGCPEARSPALTSIVSAVVG 660
QY 661 ILLVVVLGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
DB 661 ILLVVVLGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
QY 721 RKVKVLSGAGFTYKGIWIPDGENVKIPVAIKVLRNTPSKANKELDEAYVMAGVSP 780
DB 721 RKVKVLSGAGFTYKGIWIPDGENVKIPVAIKVLRNTPSKANKELDEAYVMAGVSP 780
QY 781 YVSRLLGICLTSTVQLVTLQMPYGLLDHVRNRRGRSGDQLLWNCQIAKMSYLEDVR 840
DB 781 YVSRLLGICLTSTVQLVTLQMPYGLLDHVRNRRGRSGDQLLWNCQIAKMSYLEDVR 840
QY 841 LVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRFT 900
DB 841 LVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRFT 900

Db 841 LVHRDLARNVLKSPNHVKITDFGLARLLDIDETEVHADGGKVPKMKMALESILRRFT 900
 Qy 901 HQSDVWSYGVTVWELMTFGAKPDGIGIPAREIPDLLEKGERLPQPPICTIDVYIMVKWM 960
 Db 901 HQSDVWSYGVTVWELMTFGAKPDGIGIPAREIPDLLEKGERLPQPPICTIDVYIMVKWM 960
 Qy 961 IDECRPRFRELVSERFMRDQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGLVDA 1020
 Db 961 IDECRPRFRELVSERFMRDQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGLVDA 1020
 Qy 1021 EYLVPQGFCCPDPAAGWVHRRHRSSTSGGGDLTLGLPESEERAPRPLAPSEG 1080
 Db 1021 EYLVPQGFCCPDPAAGWVHRRHRSSTSGGGDLTLGLPESEERAPRPLAPSEG 1080
 Qy 1081 AGSDVDFDGLGMAAKGLOSLPHDPSPLQRYSEDPTVLPSETDGYVAPLTCSPQPEYV 1140
 Db 1081 AGSDVDFDGLGMAAKGLOSLPHDPSPLQRYSEDPTVLPSETDGYVAPLTCSPQPEYV 1140
 Qy 1141 NOPDVRPQPSREGPLPAARPAAGATLERAKTILSPGKNGVVKOVFAFGGAVENPEYLTPO 1200
 Db 1141 NOPDVRPQPSREGPLPAARPAAGATLERAKTILSPGKNGVVKOVFAFGGAVENPEYLTPO 1200
 Qy 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGPTAENPEYLGIDVPV 1255
 Db 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGPTAENPEYLGIDVPV 1255

RESULT 4

AAU74545
 ID AAU74545 standard; Protein; 1255 AA.

AC AAU74545;

XX 23-APR-2002 (first entry)

XX Human HER2/ (ErbB2) polypeptide.

XX Human; HER2; ErbB; epidermal growth factor receptor; receptor;
 KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
 KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
 KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
 KW glial disorder; astrocytal disorder; hypothalamic disorder;
 KW glandular disorder; macrophagal disorder; epithelial disorder;
 KW stromal disorder; blastocoealic disorder; inflammatory disorder;
 KW angiogenic disorder; immunological disorder.

XX Homo sapiens.

XX US2002001587-A1.

XX 03-JAN-2002.

XX 16-MAR-2001; 2001US-0811123.

XX 16-MAR-2000; 2000US-189844P.

XX 05-OCT-2000; 2000US-238327P.

XX (ERIC/) ERICKSON S.

XX (SCHW/) SCHWALL R.

XX (SLIW/) SLIKOWSKI M.

XX Erickson S, Schwall R, Slikowski M;

XX WPI; 2002-163686/21.

XX N-PSDB; ABK14058.

XX Treating tumour characterised by overexpression of epidermal growth
 PT factor receptor, ErbB or cancer in mammal, comprises administering
 PT anti-ErbB antibody-maytansinoid conjugate to the mammal

XX Example 3; Fig 7; 93pp; English.

XX The invention relates to treating a tumour in a mammal, where the tumour

CC is characterised by the overexpression of an epidermal growth factor
 CC receptor (ErbB) and does not respond or responds poorly, to treatment
 CC with an anti-ErbB antibody, comprising administering to the mammal an
 CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for
 CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
 CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
 CC prostate and bladder, preferably breast cancer. The breast cancer is a
 CC metastatic breast cancer or an aggressive form of metastatic breast
 CC cancer which overexpresses ErbB2. The method is also useful for treating
 CC neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal,
 CC epithelial stromal, blastocoealic, inflammatory, angiogenic and
 CC immunological disorders. This sequence represents the human HER2 (ErbB2)
 CC polypeptide of the invention.

XX Sequence 1255 AA;

Query Match 98.8%; Score 6720; DB 23; Length 1255;

Best Local Similarity 99.0%; Pred. No. 0;

Matches 1242; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLLALLPPGAASDVCTGTDMLRLPASPEHLDMLRLHYQGCVVQGNL 60

Db 1 MELAALCRWGLLLALLPPGAASDVCTGTDMLRLPASPEHLDMLRLHYQGCVVQGNL 60

Qy 61 ELYLPTNASLSFLQDIQEVQGVYLIHNVQVPLQRLIRVGTOLFDENYALAVLDNG 120

Db 61 ELYLPTNASLSFLQDIQEVQGVYLIHNVQVPLQRLIRVGTOLFDENYALAVLDNG 120

Qy 121 DPLNNTPTVGTASPGGLRELQSLTEILKGGVLTORNPOLCYODTILKNDIFHKNNOLA 180

Db 121 DPLNNTPTVGTASPGGLRELQSLTEILKGGVLTORNPOLCYODTILKNDIFHKNNOLA 180

Qy 181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVTCAGGCARCKGKPLPDDCCEQC 240

Db 181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVTCAGGCARCKGKPLPDDCCEQC 240

Qy 241 AAGCTGPKHSIDCLAHFNHSGICELHCPALVQYIKANSKFITITLRYTFGASCVTAC 300

Db 241 AAGCTGPKHSIDCLAHFNHSGICELHCPALVQYIKANSKFITITLRYTFGASCVTAC 300

Qy 301 YNVLSTDVGSCTLVCPHNOEVTAEQTCCKSKPCARVCYGLGMEHLREVRVTSAN 360

Db 301 YNVLSTDVGSCTLVCPHNOEVTAEQTCCKSKPCARVCYGLGMEHLREVRVTSAN 360

Qy 361 IQEFAGCKKIFGSLAPLPESFDGDPASNTAPLOPEQLQVFETLEETITGLYISAWPDSLP 420

Db 361 IQEFAGCKKIFGSLAPLPESFDGDPASNTAPLOPEQLQVFETLEETITGLYISAWPDSLP 420

Qy 421 DLSVFQNLQVIRGRILHNGAYSILTLOGLGISWLGRLSRLRELGSGLALIHNNTHLCFVHTV 480

Db 421 DLSVFQNLQVIRGRILHNGAYSILTLOGLGISWLGRLSRLRELGSGLALIHNNTHLCFVHTV 480

Qy 481 PWDQLFRNPHOALLHTANRPEDECYGEGLACHQLCARGHCWGPGPTCCNCQSLRGQEC 540

Db 481 PWDQLFRNPHOALLHTANRPEDECYGEGLACHQLCARGHCWGPGPTCCNCQSLRGQEC 540

Qy 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPFCVAC 600

Db 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPFCVAC 600

Qy 601 PSGVKPDLISYMPIWKFPDDEGACQPCPINCTHSCVDLDDKGPAPQASPLTSIVSAVVG 660

Db 601 PSGVKPDLISYMPIWKFPDDEGACQPCPINCTHSCVDLDDKGPAPQASPLTSIVSAVVG 660

Qy 661 ILLVVLGVVFGILIKRROOKIRKVTMRLLQETELVEPLTPSGAMPNQAQRILKETEL 720

Db 661 ILLVVLGVVFGILIKRROOKIRKVTMRLLQETELVEPLTPSGAMPNQAQRILKETEL 720

Qy 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMAGVGP 780

Db 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMAGVGP 780

Qy 781 YVSRLLIGICLTSTVQLVTLQMPYGCILLDHVRENRRGLSGDQLLNWCMQIAGKMSYLEDVR 840

```
Db 781 YVSRLLGICLTSTVQLVTLMPYGLDHDVHNRGRGLSGDQLLNWCQIAKMSYLEVDV 840
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLADIDETEHADGGKVPKIMMALESILRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLADIDETEHADGGKVPKIMMALESILRRFT 900
Qy 901 HQSDVMSYGVTVWELMTFGAKPDGIPAREIPDLLEKGERLPPOPICTIDVYIMVKWM 960
Db 901 HQSDVMSYGVTVWELMTFGAKPDGIPAREIPDLLEKGERLPPOPICTIDVYIMVKWM 960
Qy 961 IDSECRPRRELVSFSEFMRDPOFVVIQNEIDLGPASPLDSTFYRSLLEDGDLVDA 1020
Db 961 IDSECRPRRELVSFSEFMRDPOFVVIQNEIDLGPASPLDSTFYRSLLEDGDLVDA 1020
Qy 1021 EYLVPQGFPCPDPAAGAGVHHRSSSTRSGGDLTLGLEPSEEAAPRPLAPSEG 1080
Db 1021 EYLVPQGFPCPDPAAGAGVHHRSSSTRSGGDLTLGLEPSEEAAPRPLAPSEG 1080
Qy 1081 AGSDVFDGDLGMAKGLQLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAKGLQLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Qy 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVAFGGAVENPEYLTQ 1200
Db 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVAFGGAVENPEYLTQ 1200
Qy 1201 GGAAPQHPHPPAPSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVPV 1255
Db 1201 GGAAPQHPHPPAPSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVPV 1255
```

RESULT 5

AAW01111 standard; Protein; 1255 AA.

AC AAW01111;

01-JAN-1997 (first entry)

DE HER-2/neu protein.

XX HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
KW breast cancer; ovary cancer; colon cancer; lung cancer;
KW prostate cancer; immunisation; tumour; vaccine; vector.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Domain 676..1255

FT /label= Intracellular domain

FT /note= "claimed domain, useful for immunisation"

XX M9963051A=NA.

XX 08-031-1936.

XX 28-MAR-1996; 96WO-US01689.

XX 31-MAR-1995; 95US-0414417.

XX (UNIW) UNIV WASHINGTON.

XX Cheever MA, Disis ML;

XX WPI; 1996-455361/45.

XX N-PSDB; AAT40739.

XX DNA encoding HER-2-neu poly.peptide(s) - used for prevention or
PT treatment of malignancies with which the HER-2/neu oncogene is
PT associated

XX Claim 2; Page 56-61; 71pp; English.

XX Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is
CC the product of the HER-2/neu oncogene (see also AAT40739). The
CC protein is over-expressed in various cancers, including breast,
CC ovarian, colon, lung and prostate. The intracellular domain of the
CC protein can be used to immunise an animal against a malignancy with
CC which the oncogene is associated. The polypeptide can be produced
CC in transfected host cells for use in immunisation. Alternatively,
CC animal cells are transfected in vivo or ex vivo with a viral vector
CC that directs expression of the polypeptide.

XX SQ Sequence 1255 AA;

Query Match 98.7%; Score 6714; DB 17; Length 1255;

Best Local Similarity 98.8%; Pred. No. 0;

Matches 1240; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLLALLPPGAASTQVCTGTDMLRLPASPTHLDMLRHLYQGCVVQGNL 60

Db 1 MELAALCRWGLLLALLPPGAASTQVCTGTDMLRLPASPTHLDMLRHLYQGCVVQGNL 60

Qy 61 ELTYLPTNASLSFLQDIQEVQYVLIHNVQVPLQRLRIVRGTLQFEDNYALAVLDNG 120

Db 61 ELTYLPTNASLSFLQDIQEVQYVLIHNVQVPLQRLRIVRGTLQFEDNYALAVLDNG 120

Qy 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQVLPOLCYQDTILWKDIIFHKNOLA 180

Db 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQVLPOLCYQDTILWKDIIFHKNOLA 180

Qy 181 LTLIDNTRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVTCAGCARCKGPLPTDCHEQC 240

Db 181 LTLIDNTRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVTCAGCARCKGPLPTDCHEQC 240

Qy 241 AAGCTGPKSDCLACLFHNSGICELHCPALVQYIKANSKFCITELRTTFGASCVTACP 300

Db 241 AAGCTGPKSDCLACLFHNSGICELHCPALVQYIKANSKFCITELRTTFGASCVTACP 300

Qy 301 YNYLSTDVSGSCTLVCPHMQEVTAEQTCCKSPKPCARVCYGLGMEHLREVRVTSAN 360

Db 301 YNYLSTDVSGSCTLVCPHMQEVTAEQTCCKSPKPCARVCYGLGMEHLREVRVTSAN 360

Qy 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFTELEEITGYLISAWPDSL 420

Db 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFTELEEITGYLISAWPDSL 420

Qy 421 DLSVFQNLQVIRGRILHNGAYSLLTQGLGISWLGSLRELGSGLALIIHHNTHLCFVHTV 480

Db 421 DLSVFQNLQVIRGRILHNGAYSLLTQGLGISWLGSLRELGSGLALIIHHNTHLCFVHTV 480

Qy 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGPTQCVCNCSQFLRGQEC 540

Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGPTQCVCNCSQFLRGQEC 540

Qy 541 VEECRVLOGLPREYVNAHCLPCHPECQPNQSVTCFGEADQCVACAHYKDPFPCVARC 600

Db 541 VEECRVLOGLPREYVNAHCLPCHPECQPNQSVTCFGEADQCVACAHYKDPFPCVARC 600

Qy 601 PSGVKPDLSPYMPIWKFPDEGACQPCPINCTHSCVDLDDKGCPEORASPLTSIVSAVWG 660

Db 601 PSGVKPDLSPYMPIWKFPDEGACQPCPINCTHSCVDLDDKGCPEORASPLTSIVSAVWG 660

Qy 661 ILLVVVLGVVFGILIKRROOKIRKYTMRLLOTELVEPLTPSGAMPNQAOMRILKETEL 720

Db 661 ILLVVVLGVVFGILIKRROOKIRKYTMRLLOTELVEPLTPSGAMPNQAOMRILKETEL 720

Qy 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVSP 780

Db 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVSP 780

Qy 781 YVSRLLGICLTSTVQLVTLMPYGLDHDVHNRGRGLSGDQLLNWCQIAKMSYLEVDV 840

Db 781 YVSRLLGICLTSTVQLVTLMPYGLDHDVHNRGRGLSGDQLLNWCQIAKMSYLEVDV 840

QY 841 LVHRDLAARNVLKSPNHHVKITDFGLARLLDIDETEHADGGKVPKWKMALESILRRFT 900
Db 841 LVHRDLAARNVLKSPNHHVKITDFGLARLLDIDETEHADGGKVPKWKMALESILRRFT 900
QY 901 HQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPPOPICTIDVYMWKCM 960
Db 901 HQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPPOPICTIDVYMWKCM 960
QY 961 IDSECRPRFRELVSERFMRDQRFVVIQNEDLGPASPLDSTFYRSLLEDDDDGDLVDA 1020
Db 961 IDSECRPRFRELVSERFMRDQRFVVIQNEDLGPASPLDSTFYRSLLEDDDDGDLVDA 1020
QY 1021 EYLVVQCGFCFDDPAPAGAGMHHRRSSSTRSGGDLTLGLEPSEEEAPRPLAPSEG 1080
Db 1021 EYLVVQCGFCFDDPAPAGAGMHHRRSSSTRSGGDLTLGLEPSEEEAPRPLAPSEG 1080
QY 1081 AGSDVFDGDLGMAAGLQSLPHDPSPLORYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAAGLQSLPHDPSPLORYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
QY 1141 NQDVRPQPSPREGPLPAARPAAGATILERAKTILSPGKNGVYKDVFAFGGAVENPEYLTPO 1200
Db 1141 NQDVRPQPSPREGPLPAARPAAGATILERAKTILSPGKNGVYKDVFAFGGAVENPEYLTPO 1200
QY 1201 GGAAPQHPHPPAFSPAFDNLVYWDQPPPERGAPPSTFKGTPPTAENPEYLGLDVPV 1255
Db 1201 GGAAPQHPHPPAFSPAFDNLVYWDQPPPERGAPPSTFKGTPPTAENPEYLGLDVPV 1255

RESULT 6

AAW92406
ID AAW92406 standard; Protein; 1255 AA.

XX AAW92406;

XX 21-APR-1999 (first entry)

DE Human HER-2/neu oncogene protein.

XX HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;
KW malignancy; treatment; tumour.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Region 676..1255

FT /note= "region which elicits immune response"

XX US5869445-A.

XX 09-FEB-1999.

XX 01-APR-1996; 96US-0625101.

XX 01-APR-1996; 96US-0625101.

PR 17-MAR-1993; 93US-0033644.

PR 12-AUG-1993; 93US-0106112.

PR 31-MAR-1995; 95US-0414417.

XX (UNIW) UNIV WASHINGTON.

XX Cheever MA, Disis ML;

XX WPI; 1999-152835/13.

XX N-PSDB; AAX01912.

XX Use of HER-2/neu polypeptides - for eliciting an immune response to
PT an HER-2/neu associated malignancy, particularly for treating or
PT preventing tumours

XX Claim 3; Column 31-38; 26pp; English.

XX This sequence represents the human HER-2/neu oncogene protein. A fragment

CC of this protein is used in a method for eliciting or enhancing an immune
CC response to HER-2/neu protein. The polypeptide can stimulate T cells and
CC B cells to produce an immune response to the HER-2/neu protein. The
CC method can be used for immunisation against a malignancy in which the
CC HER-2/neu oncogene is associated and in the treatment of an existing
CC tumour, or to prevent tumour occurrence or reoccurrence.

XX Sequence 1255 AA;

Query Match 98.7%; Score 6714; DB 20; Length 1255;

Best Local Similarity 98.8%; Pred. No. 0;

Matches 1240; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLLALLPPGAASCTGCTDMKRLRPASPETHLDMLRLHYQGCVVQGNL 60

Db 1 MELAALCRWGLLLALLPPGAASCTGCTDMKRLRPASPETHLDMLRLHYQGCVVQGNL 60

QY 61 ELYLPTNASLSFLQDIQEVQGVLIHQNVRQVPLQRLIRVGTOLFDENYALAVLDNG 120

Db 61 ELYLPTNASLSFLQDIQEVQGVLIHQNVRQVPLQRLIRVGTOLFDENYALAVLDNG 120

QY 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHKKNOLA 180

Db 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHKKNOLA 180

QY 181 LTLIDTNRSFACHPCSPMCKSGRCWGESSEDCQSLTRITVCAGGCARCKGPLPTDCCHEQC 240

Db 181 LTLIDTNRSFACHPCSPMCKSGRCWGESSEDCQSLTRITVCAGGCARCKGPLPTDCCHEQC 240

QY 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVQYIKANSKFITGIBELRTVFGASCVTACP 300

Db 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVQYIKANSKFITGIBELRTVFGASCVTACP 300

QY 301 YNYLSTDVGSCITLVCPHMQEVTAEQGTORCEKSKPCARVCYGLGMEHLREVRAVTSAN 360

Db 301 YNYLSTDVGSCITLVCPHMQEVTAEQGTORCEKSKPCARVCYGLGMEHLREVRAVTSAN 360

QY 361 IOEFAGCKKIFGSLAFPLPESFDGDPASNTAPLQPEQLQVFETLEETIGLYISAWPDSLP 420

Db 361 IOEFAGCKKIFGSLAFPLPESFDGDPASNTAPLQPEQLQVFETLEETIGLYISAWPDSLP 420

QY 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGISWGLRSLRELGSGLALIIHNTHLCFVHTV 480

Db 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGISWGLRSLRELGSGLALIIHNTHLCFVHTV 480

QY 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHOLCARGHCWGPQTQVCNCSQFLRGQEC 540

Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHOLCARGHCWGPQTQVCNCSQFLRGQEC 540

QY 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFFCVARC 600

Db 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFFCVARC 600

QY 601 PSGVKPDLISYMPIWKFPPDEBEGACQPCPNCTHSCVDLDDKGPAPORASPLTSIVSAVVG 660

Db 601 PSGVKPDLISYMPIWKFPPDEBEGACQPCPNCTHSCVDLDDKGPAPORASPLTSIVSAVVG 660

QY 661 ILLVVLGVVFGILIKRROOKIRKVTMRLLQETELVEPLTPSGAMPNQAQRILKETEL 720

Db 661 ILLVVLGVVFGILIKRROOKIRKVTMRLLQETELVEPLTPSGAMPNQAQRILKETEL 720

QY 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMAGVGSF 780

Db 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMAGVGSF 780

QY 781 YVSRLLGICLTSTVOLVTQMPYGCILLDHRNRRGLSGQDILLNMCQIAKGSYLEDVYR 840

Db 781 YVSRLLGICLTSTVOLVTQMPYGCILLDHRNRRGLSGQDILLNMCQIAKGSYLEDVYR 840

QY 841 LVHRDLAARNVLKSPNHHVKITDFGLARLLDIDETEHADGGKVPKWKMALESILRRFT 900

Db 841 LVHRDLAARNVLKSPNHHVKITDFGLARLLDIDETEHADGGKVPKWKMALESILRRFT 900

QY 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPPOPICTIDVYMWKCM 960
DB |||||
QY 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPPOPICTIDVYMWKCM 960
DB |||||
QY 961 IDSECRPRFRELVSFMRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLDDMDGLVDA 1020
DB |||||
QY 961 IDSECRPRFRELVSFMRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLDDMDGLVDA 1020
DB |||||
QY 1021 EYLVQOQFCDPAPGAGCMVHRRSSSTRSGGDLTLGLEPSEERAPRPLASEG 1080
DB |||||
QY 1021 EYLVQOQFCDPAPGAGCMVHRRSSSTRSGGDLTLGLEPSEERAPRPLASEG 1080
DB |||||
QY 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSTGTVAPLTCSPQPEYV 1140
DB |||||
QY 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSTGTVAPLTCSPQPEYV 1140
DB |||||
QY 1141 NQDVRPQPSPREGPLPAARPAAGATLERPKTILSPGKNGVVDVFAFGGAVENPEYLTPO 1200
DB |||||
QY 1141 NQDVRPQPSPREGPLPAARPAAGATLERPKTILSPGKNGVVDVFAFGGAVENPEYLTPO 1200
DB |||||
QY 1201 GGAAPQHPPPAPSPAFDNLVYWDQDPPPERGAPSTFKGTPTAENPEYLGLDVVP 1255
DB |||||
QY 1201 GGAAPQHPPPAPSPAFDNLVYWDQDPPPERGAPSTFKGTPTAENPEYLGLDVVP 1255
DB |||||

RESULT 7

ID AAB21198 standard; protein; 1255 AA.

XX AAB21198;

DT 12-JAN-2001 (first entry)
XX Human HER-2/neu protein.

DE Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;
KW breast cancer; prostate cancer; ovarian cancer; lung cancer;
KW colon cancer.

XX Homo sapiens.

XX WO200044899-A1.

XX 03-AUG-2000.

XX 28-JAN-2000; 2000WO-US02164.

XX 29-JAN-1999; 99US-0117976.

XX (CORI-) CORIXA CORP.

XX (SMIK) SMITHKLINE BEECHAM.

XX Cheever MA, Gheysen D;

XX WPI; 2000-505976/45.

XX N-PSDB; AAB89736.

XX HER-2/neu extracellular domain/phosphorylation domain fusion proteins
PT useful for vaccinating against breast, ovarian, colon, lung and
PT prostate cancers -

XX Claim 52; Fig 7; 128pp; English.

XX The present sequence is the human HER-2/neu protein. It is a member of
CC the tyrosine kinase family of receptor-like glycoproteins and shows
CC homology to the epidermal growth factor receptor (EGFR). It probably
CC plays a part in cell growth and/or differentiation. The HER-2/neu
CC gene is an oncogene. An HER-2/neu fusion protein comprising a
CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
CC domain may be used to treat or prevent cancer by eliciting or
CC enhancing an immune response to the HER-2/neu protein. It may be used
CC to treat malignancies such as breast, ovarian, colon, lung and
CC prostate cancers, and may be used as an antigen to vaccinate against

CC these neoplasias.
XX Sequence 1255 AA;
SQ Query Match 98.7%; Score 6714; DB 21; Length 1255;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 3; Mismatches 12; Indels 0; Gaps 0;
QY 1 MELAALCRWGLLALLPPGAASSTVCTGDMKRLPASPETHLDMLRHLYQCGQVQGNL 60
DB |||||
QY 1 MELAALCRWGLLALLPPGAASSTVCTGDMKRLPASPETHLDMLRHLYQCGQVQGNL 60
DB |||||
QY 61 ELTYLPTNASLSFLQDIQEVQGVLIHMQVPLQRLIRVGTOLFDNDTALAVLDNG 120
DB |||||
QY 61 ELTYLPTNASLSFLQDIQEVQGVLIHMQVPLQRLIRVGTOLFDNDTALAVLDNG 120
DB |||||
QY 121 DPLNNTTPTVGTASPGGLRELRLSLEILKGGVLIQORNPOLCYQDTILWKDIFHKKNOLA 180
DB |||||
QY 181 LTLIDTNRSRACHPCSPCKSGRCWGESSEDCQSLTRTVCCAGCARCKGPLPTDCCHEOC 240
DB |||||
QY 181 LTLIDTNRSRACHPCSPCKSGRCWGESSEDCQSLTRTVCCAGCARCKGPLPTDCCHEOC 240
DB |||||
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVQYIKANSKFITGITELRITFFGASCVTACP 300
DB |||||
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVQYIKANSKFITGITELRITFFGASCVTACP 300
DB |||||
QY 301 YNYLSTDVGSCTLVCPHNOEVTAEQTCRCKSKPCARVCYGLGMEHLREVRVTSAN 360
DB |||||
QY 301 YNYLSTDVGSCTLVCPHNOEVTAEQTCRCKSKPCARVCYGLGMEHLREVRVTSAN 360
DB |||||
QY 361 IOEFAGCKKIFGSLAPLPESPDGDPASNTAPLOEQLVFETLEETITGLYISAWPDSLP 420
DB |||||
QY 361 IOEFAGCKKIFGSLAPLPESPDGDPASNTAPLOEQLVFETLEETITGLYISAWPDSLP 420
DB |||||
QY 421 DLSVFONLQVIRGRILHNGAYSLTQGLGSLWGLRLSRELGLALIHNNHLCFVHTV 480
DB |||||
QY 421 DLSVFONLQVIRGRILHNGAYSLTQGLGSLWGLRLSRELGLALIHNNHLCFVHTV 480
DB |||||
QY 481 PWDQLFRNPQALLHTANPEDECVGEGACHQLCARGHCWGPGPTQCVNCSOFLRGQEC 540
DB |||||
QY 481 PWDQLFRNPQALLHTANPEDECVGEGACHQLCARGHCWGPGPTQCVNCSOFLRGQEC 540
DB |||||
QY 541 VEECRVLQGLPREYNNARHCLPCHPECQONGSVTCFGEADQCVACAHYKOPFFCVARC 600
DB |||||
QY 541 VEECRVLQGLPREYNNARHCLPCHPECQONGSVTCFGEADQCVACAHYKOPFFCVARC 600
DB |||||
QY 601 PSGVKPDLSPYMPKPPDEEGACQPCPINCTHSCVDLDDKGPAPQASPLTSIVSAVVG 660
DB |||||
QY 601 PSGVKPDLSPYMPKPPDEEGACQPCPINCTHSCVDLDDKGPAPQASPLTSIVSAVVG 660
DB |||||
QY 661 ILLVVVLGVVFGILIKRROOKIRKYTMRELLQETELVEPLTPSGAMPNOAQMRLKETEL 720
DB |||||
QY 661 ILLVVVLGVVFGILIKRROOKIRKYTMRELLQETELVEPLTPSGAMPNOAQMRLKETEL 720
DB |||||
QY 721 RKVKVLGSAFAGTVYKGIWIPGENVKIPVAIKVIRENTSPKANKEILDEAYVMAGVGP 780
DB |||||
QY 721 RKVKVLGSAFAGTVYKGIWIPGENVKIPVAIKVIRENTSPKANKEILDEAYVMAGVGP 780
DB |||||
QY 781 YVSRLLGICLTSTVQLVTLQMPYGLLDHVRNRRGLSGDILLNWCQAKGMSYLEDVR 840
DB |||||
QY 781 YVSRLLGICLTSTVQLVTLQMPYGLLDHVRNRRGLSGDILLNWCQAKGMSYLEDVR 840
DB |||||
QY 841 LVHRDLAARNVLKSPNNHVKITDFGLARLLDIDETEHADGGKVPKMAWLESILRRRFT 900
DB |||||
QY 841 LVHRDLAARNVLKSPNNHVKITDFGLARLLDIDETEHADGGKVPKMAWLESILRRRFT 900
DB |||||
QY 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPPOPICTIDVYMWKCM 960
DB |||||
QY 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPPOPICTIDVYMWKCM 960
DB |||||
QY 961 IDSECRPRFRELVSFMRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLDDMDGLVDA 1020
DB |||||

Db 961 IDSECRPRFRELVSFARMARDQRFVVIQNEBGLPASPJDSTFYRSLLEDDMDGLVDA 1020
QY 1021 EYLVPQOQFFCPDPAPGAGMWHHRSSSTRSGGDLTLGLEPSEEEAPRPLAPSEG 1080
Db 1021 EYLVPQOQFFCPDPAPGAGMWHHRSSSTRSGGDLTLGLEPSEEEAPRPLAPSEG 1080
QY 1081 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVAPLTCSPQPEYV 1140
QY 1141 NOPDVQPQPSREGPLPAARPAATLERAKTILSPGKNGVVVDVFAFGGAVENPEYLTPO 1200
Db 1141 NOPDVQPQPSREGPLPAARPAATLERAKTILSPGKNGVVVDVFAFGGAVENPEYLTPO 1200
QY 1201 GGAAPQHPFPAFPAFDNLYYWDQPPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
Db 1201 GGAAPQHPFPAFPAFDNLYYWDQPPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255

RESULT 8

AAy84780

ID AAY84780 standard; Protein; 1255 AA.

XX

AC AAY84780;

XX

DT 08-AUG-2000 (first entry)

XX

DE Amino acid sequence of the SPLICE erbB-2 receptor protein.

XX

KW SPLICE erbB-2 receptor protein; cell transformation disorder; cancer;
KW tumor cell proliferation; tissue degeneration; arthropathy;
KW bone resorption; inflammatory disease; degenerative disorder;
KW wound healing.

XX

OS Homo sapiens.

XX

PN W0200020579-A1.

XX

PD 13-APR-2000.

XX

PF 01-OCT-1999; 99WO-CA00912.

XX

PR 02-OCT-1998; 98US-0165192.

XX

PA (UYMC-) UNIV MCMASTER.

XX

PI Muller WJ, Siegel PM;

XX

DR WPI; 2000-303768/26.

XX

DR N-PSDB; AAA14812.

XX

PT Nucleic acid encoding an erbB 2 receptor protein designated SPLICE
PT erbB-2, inhibitors of the protein are useful for treatment of cancer -
PS Claim 3; Fig 2; 60pp; English.

XX

The present sequence represents a SPLICE erbB-2 receptor protein. The protein has an in-frame deletion of 16 amino acids, 2 of which are conserved cysteine residues, compared to the unspliced protein. The erbB-2 polynucleotide is used to construct probes for detecting disorders of cell transformation such as cancer. Antibodies to the protein may be used to detect SPLICE erbB-2 in a sample. Agents (e.g. antisense oligonucleotides) which inhibit the expression of SPLICE erbB-2 are useful for reducing tumor cell proliferation and treating cancer. Substances which stimulate SPLICE erbB-2 are useful for treating conditions involving damaged cells including conditions in which degeneration of tissue occurs, such as arthropathy, bone resorption, inflammatory diseases, degenerative disorders of the central nervous system and wound healing.

XX

SQ Sequence 1255 AA;

XX

QY 1021 EYLVPQOQFFCPDPAPGAGMWHHRSSSTRSGGDLTLGLEPSEEEAPRPLAPSEG 1080

Query Match 98.7%; Score 6714; DB 21; Length 1255;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 1 MELAALCRWELLALLPPCAASTOVCCTGDMKLRIPASPETHLDMLRLHYQGCVQVGNL 60
Db 1 MELAALCRWELLALLPPCAASTOVCCTGDMKLRIPASPETHLDMLRLHYQGCVQVGNL 60
QY 61 ELTYLTPTNASLSFLQDIQEVQGVLIHNVQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
Db 61 ELTYLTPTNASLSFLQDIQEVQGVLIHNVQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
QY 121 DPLNNTPTVTGASPGGLRELQRLSTEILKGGVLIQORNPOLCYQDTILKWDIFHKNNQLA 180
Db 121 DPLNNTPTVTGASPGGLRELQRLSTEILKGGVLIQORNPOLCYQDTILKWDIFHKNNQLA 180
QY 181 LTLIDTNRGRACHPCSPCKGSRGWESSEDCQSLTRTVACGACARCKGPLTDCCHEQC 240
Db 181 LTLIDTNRGRACHPCSPCKGSRGWESSEDCQSLTRTVACGACARCKGPLTDCCHEQC 240
QY 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVQYIKANSKFICITELRYTTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVYNTDTFESMPNPSEGYTTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTORCEKCKPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTORCEKCKPCARVCYGLGMEHLREVRVTSAN 360
QY 361 IOEFAGCKKIFGSLAPLPESFDGDPASNTAPLOPOLQVTFLEETITGVLYISAWPDSLP 420
Db 361 IOEFAGCKKIFGSLAPLPESFDGDPASNTAPLOPOLQVTFLEETITGVLYISAWPDSLP 420
QY 421 DLSVFONLQVIRGRILHNGAYSILTQGLGISWGLSRLSRELGSGLALIHNNHLCFVHTV 480
Db 421 DLSVFONLQVIRGRILHNGAYSILTQGLGISWGLSRLSRELGSGLALIHNNHLCFVHTV 480
QY 481 PWDQLFRNPHQALLHTANRPEDECVGEGIACHQLCARGHCWGPPTQCVNCSQFLRGQEC 540
Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGIACHQLCARGHCWGPPTQCVNCSQFLRGQEC 540
QY 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKOPPCVARC 600
Db 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKOPPCVARC 600
QY 601 PSGVKPDLSPYMPDKFPPDEGACQPCPINCTHSCVDLDDKGCAPAEORASPLTSIVSAVVG 660
Db 601 PSGVKPDLSPYMPDKFPPDEGACQPCPINCTHSCVDLDDKGCAPAEORASPLTSIVSAVVG 660
QY 661 ILLVVVLGVVFGILIKRRQOKIRKYTMRLLOTELVEPLTPSGAMPNQAQMRILKETEL 720
Db 661 ILLVVVLGVVFGILIKRRQOKIRKYTMRLLOTELVEPLTPSGAMPNQAQMRILKETEL 720
QY 721 RKVKVLGSAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYMAGVGP 780
Db 721 RKVKVLGSAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYMAGVGP 780
QY 781 YVSRLLGICLTSTVOLVTOLMPYGLLDHVRNRRGLSGDQLLNCWQIAGMSYLEDVYR 840
Db 781 YVSRLLGICLTSTVOLVTOLMPYGLLDHVRNRRGLSGDQLLNCWQIAGMSYLEDVYR 840
QY 841 LVHRDLAARNVLKSPNVHKITDFGLARLLDIDETEHADGGKVPKWMMALESILRRRFT 900
Db 841 LVHRDLAARNVLKSPNVHKITDFGLARLLDIDETEHADGGKVPKWMMALESILRRRFT 900
QY 901 HQSDVMSYGVTVWELMTFGAKYDIPAREIPDLLEKGERLPPQPICTIDVTMVMKCM 960
Db 901 HQSDVMSYGVTVWELMTFGAKYDIPAREIPDLLEKGERLPPQPICTIDVTMVMKCM 960
QY 961 IDSECRPRFRELVSFARMARDQRFVVIQNEBGLPASPJDSTFYRSLLEDDMDGLVDA 1020
Db 961 IDSECRPRFRELVSFARMARDQRFVVIQNEBGLPASPJDSTFYRSLLEDDMDGLVDA 1020

Db 1021 EEVLVPQQGFCDPAPACAGMHHRRSSSTRSGGDLTLGLEPSEEAAPRSPAPSEG 1080
QY 1081 AGSDVFDGDLGMAKGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAKGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
QY 1141 NQPDVRQPSPREGPLPAARPAAGATLERAKTLPSPGKGVVGVKDVAFGGGAVENPEYLTPO 1200
Db 1141 NQPDVRQPSPREGPLPAARPAAGATLERAKTLPSPGKGVVGVKDVAFGGGAVENPEYLTPO 1200
QY 1201 GGAAPQHPPPAFSPADNLYMDQDPPERGAPSTFKGTPTAENPEYLGDLVVP 1255
Db 1201 GGAAPQHPPPAFSPADNLYMDQDPPERGAPSTFKGTPTAENPEYLGDLVVP 1255

RESULT 9
AAB85458
ID AAB85458 standard; Protein; 1255 AA.
XX AAB85458;
XX AC
XX DE
DT 25-SEP-2001 (first entry)
XX Human HER-2/neu protein.
XX Antigen-presenting cell; immunogenic; immune response; HER-2/neu;
KW oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.
XX Homo sapiens.
XX OS
XX WO200153463-A2.
XX PN
XX PD
XX PF
XX PR
XX PS
XX (CORI-) CORIXA CORP.
XX Cheever MA, Hand-Zimmermann S;
XX WPI; 2001-476112/51.
XX N-PSDB; AAH23392.
XX New antigen-presenting cells, useful as vaccines for eliciting or
PT enhancing an immune response to HER-2/neu protein, particularly useful
PT for treating or preventing cancer, e.g. breast cancer
XX
XX Claim 2; Page 41-46; 49pp; English.
XX
XX The invention provides an isolated antigen-presenting cell, which
CC expresses at least an immunogenic portion of a polypeptide that produces
CC an immune response to HER-2/neu protein. The antigen-presenting cells are
CC useful as vaccines for eliciting or enhancing an immune response to
CC HER-2/neu protein, particularly in treating or preventing malignancies in
CC which the HER-2/neu oncogene is associated. Specifically, these are
CC useful for treating or preventing cancer, e.g. breast cancer, ovarian,
CC colon, lung or prostate cancers. The present sequence represents
CC the human HER-2/neu protein (also known as p185 or c-erbB2).
XX
SQ Sequence 1255 AA;

Query Match 98.7%; Score 6714; DB 22; Length 1255;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 1 MELALCRWGLLALLPPGAASCTGCTDMKRLRPASPTHLDMLRHLHYQGCVVQGNL 60
Db 1 MELALCRWGLLALLPPGAASCTGCTDMKRLRPASPTHLDMLRHLHYQGCVVQGNL 60
QY 61 ELTYLPTNASLFLQDIQEVQGVYLIHQAQVPLQRLRIRVRGTQLFEDNYALAVLDNG 1200

Db 61 ELTYLPTNASLFLQDIQEVQGVYLIHQAQVPLQRLRIRVRGTQLFEDNYALAVLDNG 120
QY 121 DPLNNTTPTVGTGSPGGLRELQRLSLTEILKGGVLTORNPOLCYQDITLWKDIFPHKNOLA 180
Db 121 DPLNNTTPTVGTGSPGGLRELQRLSLTEILKGGVLTORNPOLCYQDITLWKDIFPHKNOLA 180
QY 181 LTLIDNRRACHPCSPMCKGRCWGESSEDCQSLTRTVCAAGCARCKGKPLPTDCCHEQC 240
Db 181 LTLIDNRRACHPCSPMCKGRCWGESSEDCQSLTRTVCAAGCARCKGKPLPTDCCHEQC 240
QY 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVOYIKANSKFITGILRITRYFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVOYIKANSKFITGILRITRYFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPHLHQEVTAEDGTORCEKCKPCARVCYGLGMEHLREVRATYSAN 360
Db 301 YNYLSTDVGSCTLVCPHLHQEVTAEDGTORCEKCKPCARVCYGLGMEHLREVRATYSAN 360
QY 361 IOEFAGCKKIFGSLAFPLPSFDGDPASNTAPIQLOVFEETLEETGYLYISAWPDSL 420
Db 361 IOEFAGCKKIFGSLAFPLPSFDGDPASNTAPIQLOVFEETLEETGYLYISAWPDSL 420
QY 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWGLRSLRELGLALIHHTHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWGLRSLRELGLALIHHTHLCFVHTV 480
QY 481 PWDQLFRNPHQALLHTANRPEDECVEGLACHQLCARGHCWGPGPTQCVCNCSOFLRGQC 540
Db 481 PWDQLFRNPHQALLHTANRPEDECVEGLACHQLCARGHCWGPGPTQCVCNCSOFLRGQC 540
QY 541 VEECRVLQGLPREYNARHCLPCHPCQPNQSGVTCFGEADQCVACAHYKDPFPCVARC 600
Db 541 VEECRVLQGLPREYNARHCLPCHPCQPNQSGVTCFGEADQCVACAHYKDPFPCVARC 600
QY 601 PSGVKPDLSPYMPKPPDEEGACQPCPINCTHSCVDLDDKGPAPAEQASPLTSIVSAVVG 660
Db 601 PSGVKPDLSPYMPKPPDEEGACQPCPINCTHSCVDLDDKGPAPAEQASPLTSIVSAVVG 660
QY 661 ILLVVLGVVFGILIKRROOKIKYTMRLLOETELVEPLTPSGAMPNQAOMRLKETEL 720
Db 661 ILLVVLGVVFGILIKRROOKIKYTMRLLOETELVEPLTPSGAMPNQAOMRLKETEL 720
QY 721 RKVKVLGSGAFGTVYKGIWIPDENVKI PVAIKVLRNTSPKANKEILDEAYVMAGVGP 780
Db 721 RKVKVLGSGAFGTVYKGIWIPDENVKI PVAIKVLRNTSPKANKEILDEAYVMAGVGP 780
QY 781 YVSRLLGICLTSTVQLVTQMPYGLLDHVRNRLGSDQLLNWCMIAGKMSYLEDVR 840
Db 781 YVSRLLGICLTSTVQLVTQMPYGLLDHVRNRLGSDQLLNWCMIAGKMSYLEDVR 840
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKKNWALSILRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKKNWALSILRRFT 900
QY 901 HOSDWSYGVTVWELMTFCAKPYDGI PAREIPDLLEKGERLPQPPICITDVMYMWKCM 960
Db 901 HOSDWSYGVTVWELMTFCAKPYDGI PAREIPDLLEKGERLPQPPICITDVMYMWKCM 960
QY 961 IDSECRPRFRELVSFARMARDPQRFVVIQNEIDLGSPASPLDSTFYRSLLEDDMDGLVDA 1020
Db 961 IDSECRPRFRELVSFARMARDPQRFVVIQNEIDLGSPASPLDSTFYRSLLEDDMDGLVDA 1020
QY 1021 EYLVPOQGFCDPAPAGAGMHHRRSSSTRSGGDLTLGLEPSEEAAPRSPAPSEG 1080
Db 1021 EYLVPOQGFCDPAPAGAGMHHRRSSSTRSGGDLTLGLEPSEEAAPRSPAPSEG 1080
QY 1081 AGSDVFDGDLGMAKGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAKGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
QY 1141 NQPDVRQPSPREGPLPAARPAAGATLERAKTLPSPGKGVVGVKDVAFGGGAVENPEYLTPO 1200

Db 1141 NQPDVRQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVAFGGAVENPEYLTPQ 1200
QY 1201 GGAAPQHPPPAPSPADNLYYWDQPPERGAPPSTFKGTPTAENPEYGLDVPV 1255
Db 1201 GGAAPQHPPPAPSPADNLYYWDQPPERGAPPSTFKGTPTAENPEYGLDVPV 1255

RESULT 10

AG88267
ID AAG88267 standard; Protein; 1255 AA.

XX AC AAG88267;

XX 11-SEP-2001 (first entry)

XX HER2/neu amino acid sequence.

XX Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
KW immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;
KW tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.

XX Homo sapiens.

XX WO200141787-A1.

XX 14-JUN-2001.

XX 11-DEC-2000; 2000WO-US33591.

XX 10-DEC-1999; 99US-0458299.

XX (EPIM-) EPIMUNE INC.

XX Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;

XX keogh E;

XX WPI; 2001-374995/39.

XX An isolated prepared HER2/neu epitope useful in a vaccine for inducing
PT cellular immune responses for the prevention and treatment of cancer -

XX Disclosure; Page 15; 199pp; English.

XX The present invention describes isolated prepared HER2/neu epitopes (I).
CC Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
CC culture in vitro and binds to a complex of an epitope (I), bound to a
CC human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)
CC and a second epitope and the peptide is less than 50 contiguous amino
CC acids that have 100% identity with a native peptide sequence of HER2/neu;
CC (3) a vaccine composition (III) comprising (II) and a pharmaceutical
CC excipient; (4) an isolated nucleic acid encoding a peptide comprising
CC (I); and (5) an isolated nucleic acid encoding (II). (I) has cytostatic
CC and immunostimulant activities, and can be used in vaccines. (I), (II)
CC and (III) are useful for inducing cellular immune responses for the
CC prevention and treatment of cancer. (I) and (II) are useful for
CC monitoring or evaluating an immune response to a tumour-associated
CC antigen when incubated with a T lymphocyte sample form a patient and
CC detecting the presence of bound T lymphocyte to (I) or (II). Epitope
CC based vaccines mean that immunosuppressive epitopes that may be present
CC in whole antigens may be avoided. Selected epitopes may be combined to
CC enhance immunogenicity. The possible pathological side effects caused by
CC infectious agents or whole protein antigen is eliminated. The vaccine
CC provides the ability to direct and focus an immune response to multiple
CC selected antigens from the same pathogen. Epitope-based anti-tumour
CC vaccines provides the opportunity to combine epitopes derived from
CC multiple tumour-associated molecules addressing the problem of tumour-
CC tumour variability and reducing the likelihood of tumour escape due to
CC antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
CC the exemplification of the present invention.

XX Sequence 1255 AA;

XX Query Match 98.7%; Score 6714; DB 22; Length 1255;

Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLLALLPPGAASCTGCTDMKRLRPASPETHLDMLRHLYQGCVVQGNL 60

Db 1 MELAALCRWGLLLALLPPGAASCTGCTDMKRLRPASPETHLDMLRHLYQGCVVQGNL 60

QY 61 ELTYLPTNASLSFLODIQEVQGVYLIHNOVQVPLQRLIRVGTOLFDNTALAVLDNG 120

Db 61 ELTYLPTNASLSFLODIQEVQGVYLIHNOVQVPLQRLIRVGTOLFDNTALAVLDNG 120

QY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLTORNPOLCYODTILAKDIFHKNNOLA 180

Db 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLTORNPOLCYODTILAKDIFHKNNOLA 180

QY 181 LTLIDTNRSRACHPCSPCKSGRCWGESSEDCQSUTRTVCAGGCARCKGKPLPTDCCHEQC 240

Db 181 LTLIDTNRSRACHPCSPCKSGRCWGESSEDCQSUTRTVCAGGCARCKGKPLPTDCCHEQC 240

QY 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVQYIKANSKFIGITELRYTFGASCVTACP 300

Db 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVQYIKANSKFIGITELRYTFGASCVTACP 300

QY 301 YNYLSTDVGSCTLVCPHLHNOEVTAEQTCRCKSPCARVCYGLGMEHLREVRAVTSAN 360

Db 301 YNYLSTDVGSCTLVCPHLHNOEVTAEQTCRCKSPCARVCYGLGMEHLREVRAVTSAN 360

QY 361 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOVFEETLEETITGLYLSAWPDSLP 420

Db 361 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOVFEETLEETITGLYLSAWPDSLP 420

QY 421 DLSVFQNLQVIRGRILHNGAYSLTIQGLGISWGLRSLRELGSGLALIHNNHLCFVHTV 480

Db 421 DLSVFQNLQVIRGRILHNGAYSLTIQGLGISWGLRSLRELGSGLALIHNNHLCFVHTV 480

QY 481 PWDQLFRPHQALHTANRPEDECYVGEGLACHQLCARGHCWGPPTQVCNCSQFLRGQBC 540

Db 481 PWDQLFRPHQALHTANRPEDECYVGEGLACHQLCARGHCWGPPTQVCNCSQFLRGQBC 540

QY 541 VVEECRVLOQLPREYVNAHCLPCHPECQPNQSVTCFGEADQCVACAHYKDPFPCVAC 600

Db 541 VVEECRVLOQLPREYVNAHCLPCHPECQPNQSVTCFGEADQCVACAHYKDPFPCVAC 600

QY 601 PSGVKPDLSYMPIMKFPDEEGACQPCINCHTHSCVDLDDKGPACQASPLTSISAVVG 660

Db 601 PSGVKPDLSYMPIMKFPDEEGACQPCINCHTHSCVDLDDKGPACQASPLTSISAVVG 660

QY 661 ILLVVVLGVVFGILIKRQOKIRKYMRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720

Db 661 ILLVVVLGVVFGILIKRQOKIRKYMRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720

QY 721 RKVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMAGVGP 780

Db 721 RKVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMAGVGP 780

QY 781 YVSRLLGICLTSTVOLVTQLMPEYGLLDHVRNRRGLSQDILLNWCQIAKMSYLEDVYR 840

Db 781 YVSRLLGICLTSTVOLVTQLMPEYGLLDHVRNRRGLSQDILLNWCQIAKMSYLEDVYR 840

QY 841 LVHRDLAARNVLKSPNHNKVTDFGLARLLDIDETEHADGKVPKKNWALSILRRRRT 900

Db 841 LVHRDLAARNVLKSPNHNKVTDFGLARLLDIDETEHADGKVPKKNWALSILRRRRT 900

QY 901 HOSDVWSVGVTVWEIMLTCAKPYDGI PAREIPDLLEKGERLPOPPICITDVYMWKWM 960

Db 901 HOSDVWSVGVTVWEIMLTCAKPYDGI PAREIPDLLEKGERLPOPPICITDVYMWKWM 960

QY 961 IDSECEPRFRELVSFSSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMDGLVDA 1020

Db 961 IDSECEPRFRELVSFSSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMDGLVDA 1020

QY 1021 EBYLVPQGFCCPDPAAGAGVMVHRSSSTRSGGDLTLGLEPSEBEPASPLAPSEG 1080

Db 1021 EBYLVPQGFCCPDPAAGAGVMVHRSSSTRSGGDLTLGLEPSEBEPASPLAPSEG 1080

Db 1021 EYLVPQQFFCPDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
 Qy 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSDPTVLPSETDGYVAPLTCSQPQPEYV 1140
 Db 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSDPTVLPSETDGYVAPLTCSQPQPEYV 1140
 Qy 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVYKOVFAFGGAVENPEYLTQP 1200
 Db 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERPKTLSPGKNGVYKOVFAFGGAVENPEYLTQP 1200
 Qy 1201 GGAAPQHPPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPYLGLDVVPV 1255
 Db 1201 GGAAPQHPPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPYLGLDVVPV 1255

RESULT 11

AAE24067
 ID AAE24067 standard; Protein; 1255 AA.

AC AAE24067;

XX 23-SEP-2002 (first entry)

DT Human Her-2 protein.

XX Human; Her-2; epidermal growth factor receptor 2; infection; cancer;

KW hyperproliferative disorder; prophylaxis; inflammation; antisense;

KW tumour; gene therapy; phosphorothioate backbone.

XX Homo sapiens.

OS WO200222636-A1.

PN 21-MAR-2002.

PD 12-SEP-2001; 2001WO-US28572.

XX 15-SEP-2000; 2000US-0663834.

PR (ISIS-) ISIS PHARM INC.

PA Bennett CF, Cowser LM;

PI WPI; 2002-471192/50.

DR N-PSDB; AAD38904.

XX Novel antisense oligonucleotide which modulates the expression of Human

PT Epidermal Growth Factor receptor, Her2, is useful for treating tumors

PT inflammation or to prevent infection in humans -

XX Example 13; Page 95-107; 116pp; English.

XX The invention relates to antisense compounds targetted to a nucleic

CC acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)

CC that specifically hybridises with and inhibits the expression of Her2.

CC Antisense compounds of the invention are used for treating diseases or

CC conditions associated with Her2 such as hyperproliferative disorders

CC e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,

CC neural or cardiac cancer. They are also useful prophylactically e.g.

CC to prevent or delay infection, inflammation and tumour formation. The

CC invention is also used in gene therapy. The present sequence is human

CC Her-2 protein.

XX Sequence 1255 AA;

Query Match 98.7%; Score 6714; DB 23; Length 1255;
 Best Local Similarity 98.8%; Pred. No. 0;
 Matches 1240; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLALLPPGAASQVCTGTDMLRLPASPEHLDMLRHLYQGQCVVQGNL 60

Db 1 MELAALCRWGLLALLPPGAASQVCTGTDMLRLPASPEHLDMLRHLYQGQCVVQGNL 60

Qy 61 ELYLPTNASTSLFLODIEVQGYVLIHAHQVQVPLQRLRIVRGTOLEFEDNYALAVLDNG 120
 Db 61 ELYLPTNASTSLFLODIEVQGYVLIHAHQVQVPLQRLRIVRGTOLEFEDNYALAVLDNG 120
 Qy 121 DPLNNTTPTVGASPGGLRELQLASLTIELKGGVLIQORNPOLCYQDITLWKDILFHKNOLA 180
 Db 121 DPLNNTTPTVGASPGGLRELQLASLTIELKGGVLIQORNPOLCYQDITLWKDILFHKNOLA 180
 Qy 181 LTLIDTNRARACHPCSPCKGSCWGESSEDCOSLRTVTCAGGCARCKGPLPDDCCHEQC 240
 Db 181 LTLIDTNRARACHPCSPCKGSCWGESSEDCOSLRTVTCAGGCARCKGPLPDDCCHEQC 240
 Qy 241 AAGCTGPKGSDCLACLFHNFHSGICEHFCPALVYIKANSKFIGITELRYTFGASCVTACP 300
 Db 241 AAGCTGPKGSDCLACLFHNFHSGICEHFCPALVYIKANSKFIGITELRYTFGASCVTACP 300
 Qy 301 YNYLSTDVGSCTLVCPHNOEVTAEQTCRCKSPCARVCYGLGMEHLREVRVTSAN 360
 Db 301 YNYLSTDVGSCTLVCPHNOEVTAEQTCRCKSPCARVCYGLGMEHLREVRVTSAN 360
 Qy 361 IQEFAGCKITFGSLAFLPESFDGDPASNTAPLOPELOQVPELLEITGYLIISAWPDSL 420
 Db 361 IQEFAGCKITFGSLAFLPESFDGDPASNTAPLOPELOQVPELLEITGYLIISAWPDSL 420
 Qy 421 DLSVFQNLQVIRGRIILHNGAYSITLQGLGTSWLGSLRELGLSLALIHNTHLFCFVHTV 480
 Db 421 DLSVFQNLQVIRGRIILHNGAYSITLQGLGTSWLGSLRELGLSLALIHNTHLFCFVHTV 480
 Qy 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAACHOLCARGHCWGPGPTQCVNCSOFLRGQEC 540
 Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAACHOLCARGHCWGPGPTQCVNCSOFLRGQEC 540
 Qy 541 VEECRVLOGLPREYVVARHCLPCHPECOPQNGSVTCFGEADOCVCAHYKDPFCVARC 600
 Db 541 VEECRVLOGLPREYVVARHCLPCHPECOPQNGSVTCFGEADOCVCAHYKDPFCVARC 600
 Qy 601 PSGVKPDLSPYMPIWKFPPDEEGACQPCPINCTHSCVDLDDKGCPEARASPLTSIVSAVVG 660
 Db 601 PSGVKPDLSPYMPIWKFPPDEEGACQPCPINCTHSCVDLDDKGCPEARASPLTSIVSAVVG 660
 Qy 661 ILLVVVLGVVFGILIKRROOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
 Db 661 ILLVVVLGVVFGILIKRROOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
 Qy 721 RKVKVLGSGAFCTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKIILDEAYVMAGVSP 780
 Db 721 RKVKVLGSGAFCTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKIILDEAYVMAGVSP 780
 Qy 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVNRGRGLSQDILLNWCWQIAKGMYSYLEDVR 840
 Db 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVNRGRGLSQDILLNWCWQIAKGMYSYLEDVR 840
 Qy 841 LVHRDLAARNVLKSPNNHVKITDFGLARLLDIDETEHADGGKVPKMALESILRRRFT 900
 Db 841 LVHRDLAARNVLKSPNNHVKITDFGLARLLDIDETEHADGGKVPKMALESILRRRFT 900
 Qy 901 HQSDVMSYGVVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPCTIDVVMIMVKCM 960
 Db 901 HQSDVMSYGVVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPCTIDVVMIMVKCM 960
 Qy 961 IDSECRPRRELVSSEFSRMARDPQRFVJQNEIDLGPASPLDSTFYRSLLEDDDDMDGLVDA 1020
 Db 961 IDSECRPRRELVSSEFSRMARDPQRFVJQNEIDLGPASPLDSTFYRSLLEDDDDMDGLVDA 1020
 Qy 1021 EYLVPQQFFCPDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
 Db 1021 EYLVPQQFFCPDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
 Qy 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSDPTVLPSETDGYVAPLTCSQPQPEYV 1140
 Db 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSDPTVLPSETDGYVAPLTCSQPQPEYV 1140
 Qy 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVYKOVFAFGGAVENPEYLTQP 1200

Db 1141 NQDVRPQPPSPREGPLAARPAAGATLERPKTSLSPGKNGVVKDVFAPGGAVENPEYLTQP 1200
QY 1201 GGAAPQHPPPAFSPAFDNLYYWDQPPPPRGAPSPFTKGTPTAENPEYGLDVPV 1255
Db 1201 GGAAPQHPPPAFSPAFDNLYYWDQPPPPRGAPSPFTKGTPTAENPEYGLDVPV 1255
RESULT 12
AAE20479
ID AAE20479 standard; Protein; 1255 AA.
XX AAE20479;
AC AAE20479;
XX 01-JUL-2002 (first entry)
DT Human Her-2/neu protein.
XX Human; Her-2/Neu protein; immune response; gene therapy; breast cancer;
KW human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.
XX Homo sapiens.
XX Key Location/Qualifiers
FT Region 1021..1030
FT /note= "Naturally processed HLA-B44-restricted epitope"
XX WO200214503-A2.
XX 21-FEB-2002.
XX 14-AUG-2001; 2001WO-US41733.
XX 14-AUG-2000; 2000US-225152P.
XX 28-SEP-2000; 2000US-236428P.
XX 21-FEB-2001; 2001US-270520P.
XX (CORI-) CORIXA CORP.
XX Hand-zimmermann S, Cheever MA, Foy TM, Lodes MJ, Kalos MD;
PI McNeill PD, Vedvick TS;
XX WPI; 2002-280758/32.
DR N-PSDB; AAD32743.
XX Novel isolated Her-2/Neu polypeptide composition useful for therapy,
PT prevention and diagnosis of cancer, preferably breast cancer
XX Disclosure; Page 114-117; 129pp; English.
XX The invention relates to an isolated Her-2/Neu polypeptide composition
CC effective for eliciting an immune response. The invention is useful for
CC eliciting an immune response in a patient, where the patient is human
CC leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.
CC The composition is useful for the therapy and diagnosis of cancer,
CC preferably breast cancer, in pharmaceutical compositions, e.g., vaccine
CC and other compositions for the diagnosis, prevention and treatment of
CC human malignancies, for stimulating and/or expanding T cells specific for
CC Her-2/Neu polypeptide and for inhibiting the development of cancer in a
CC patient. The invention is useful for stimulating a T cell response in a
CC human patient, as probe or primer for nucleic acid hybridisation, to
CC selectively form duplex molecules with complementary stretches of the
CC entire Her-2/Neu gene or gene fragments of interest, to isolate a full
CC length gene from a suitable library, and to direct expression of a
CC polypeptide in appropriate host cells. The composition is useful in
CC prophylactic or therapeutic applications and for the treatment of cancer,
CC preferably for the immunotherapy of breast cancer and other Her-2/Neu-
CC associated malignancies. The invention is useful in gene therapy. The
CC present sequence is human Her-2/neu protein.
XX Sequence 1255 AA;
SQ

Query Match

98.7%; Score 6714; DB 23; Length 1255;

Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 3; Mismatches 12; Indels 0; Gaps 0;
QY 1 MELAALCRWGLLLALLPPGAASSTQVCTGDMKRLRLPASPEHLDMLRHLYQGCVVQGNL 60
Db 1 MELAALCRWGLLLALLPPGAASSTQVCTGDMKRLRLPASPEHLDMLRHLYQGCVVQGNL 60
QY 61 ELTYLPTNASLSFLQDIQEVQGVYLIHAHQVQVPLRLIRVRGTOLFDENYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIQEVQGVYLIHAHQVQVPLRLIRVRGTOLFDENYALAVLDNG 120
QY 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVILQRPOLCYODTILMKDIFHKKNOLA 180
Db 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVILQRPOLCYODTILMKDIFHKKNOLA 180
QY 181 LTLIDNRSFACHPCSPCKGSRCESSQSLTRTVACGACARCKGKPLPTDCCHEOC 240
Db 181 LTLIDNRSFACHPCSPCKGSRCESSQSLTRTVACGACARCKGKPLPTDCCHEOC 240
QY 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVOYIKANSKFITELRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVOYIKANSKFITELRYTFGASCVTACP 300
QY 301 YNYLSTDVGCCTLVCPLNHQVTAEDGTORCEKSKPCARVCYGLGMEHLREVRATVSAN 360
Db 301 YNYLSTDVGCCTLVCPLNHQVTAEDGTORCEKSKPCARVCYGLGMEHLREVRATVSAN 360
QY 361 IOEFAGCKKI FGSLAFLPESFDGDPASNTAPLQPELOQVFETLEETIGYLYISAWPDSL 420
Db 361 IOEFAGCKKI FGSLAFLPESFDGDPASNTAPLQPELOQVFETLEETIGYLYISAWPDSL 420
QY 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWGLRSLRELGLSLALIHNNHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWGLRSLRELGLSLALIHNNHLCFVHTV 480
QY 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQCHGWCWPGPTQCNCQFRLGQEC 540
Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQCHGWCWPGPTQCNCQFRLGQEC 540
QY 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADOCVACAHYKOPPCVAC 600
Db 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADOCVACAHYKOPPCVAC 600
QY 601 PSGVKPDLSPYMPKPFPPDEEGACQPCINCTHSCVDLDDKGPAPQASPLTSIVSAVVG 660
Db 601 PSGVKPDLSPYMPKPFPPDEEGACQPCINCTHSCVDLDDKGPAPQASPLTSIVSAVVG 660
QY 661 ILLVVLGVVFGILIKRROQKIRKVTMRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
Db 661 ILLVVLGVVFGILIKRROQKIRKVTMRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
QY 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVALKVLRENTSPKANKEILDEAYVMAGVGP 780
Db 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVALKVLRENTSPKANKEILDEAYVMAGVGP 780
QY 781 YVSRLLGICLTSTVOLVTQLMFYGLLDHVRNRRGLSGQDLLNWCWQIAGKMSYLEDVR 840
Db 781 YVSRLLGICLTSTVOLVTQLMFYGLLDHVRNRRGLSGQDLLNWCWQIAGKMSYLEDVR 840
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKKNWMALESILRRRT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKKNWMALESILRRRT 900
QY 901 HQSDVMSYGVTVWELMTFCAKEDYDGPAREIPDLLEKGERLPQPPCTIDVYMWKCM 960
Db 901 HQSDVMSYGVTVWELMTFCAKEDYDGPAREIPDLLEKGERLPQPPCTIDVYMWKCM 960
QY 961 IDSECRPRELVSEFSRWARDPQRFVITQNEDLGPASPLDSTFYRSLLDEDDMGDLVDA 1020
Db 961 IDSECRPRELVSEFSRWARDPQRFVITQNEDLGPASPLDSTFYRSLLDEDDMGDLVDA 1020
QY 1021 EBYLVPQGGFFCDDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEPAPSLAPSG 1080

Db 1021 EYLVPQQGFCPPAPAGAGWVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
 Qy 1081 AGSDVFDGLGMAKGLSLPHDPSPLQYSEDPTVPLPSSTDGVVAPLTCSPQPEYV 1140
 Db 1081 AGSDVFDGLGMAKGLSLPHDPSPLQYSEDPTVPLPSSTDGVVAPLTCSPQPEYV 1140
 Qy 1141 NQPDVRQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAGGAVENPEYLTPO 1200
 Db 1141 NQPDVRQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAGGAVENPEYLTPO 1200
 Qy 1201 GGAAPQHPAPPAPAFDNLYYWDQDPPERGAPPSTPKGPTTAENPEYLGLDVVPV 1255
 Db 1201 GGAAPQHPAPPAPAFDNLYYWDQDPPERGAPPSTPKGPTTAENPEYLGLDVVPV 1255

RESULT 13

AAM51143
 ID AAM51143 standard; Protein; 1255 AA.

XX AC AAM51143;
 XX 17-JUN-2002 (first entry)
 XX Human Her-2/neu oncogene-encoded p185 glycoprotein.
 XX Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;
 KW tyrosine kinase; receptor; c-erbB2; gene therapy.
 XX Homo sapiens.

XX Key Location/Qualifiers
 FH Domain 1..653
 FT /note= "extracellular domain"
 FT Domain 676..1255
 FT /note= "intracellular domain"
 FT Domain 990..1255
 FT /note= "phosphorylation domain"
 XX W0200212341-A2.

XX 14-FEB-2002.

XX 03-AUG-2001; 2001WO-US24283.

XX 03-AUG-2000; 2000US-0632507.

XX (CORI-) CORIXA CORP.
 PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX Cheever MA, Gheysen D;

XX WPI; 2002-241743/29.

XX N-PSDB; ABA92250.

XX Her-2/neu fusion protein for treating or preventing cancer by eliciting
 or enhancing an immune response to the protein, has Her-2/neu
 extracellular domain fused to Her-2/neu intracellular or
 phosphorylation domain

XX Claim 68; Fig 7; 141pp; English.

XX The present sequence is that of human Her-2/neu (p185 glycoprotein
 or c-erbB2), an oncogenic self-protein and target for anti-cancer
 vaccines. The Her-2/neu gene is amplified and p185 is overexpressed
 in a variety of cancers, including breast, ovarian, colon, lung and
 prostate cancer. Her-2/neu is a member of the tyrosine kinase
 family of receptor-like glycoproteins. It comprises an extracellular
 domain with homology to the epidermal growth factor receptor
 (EGFR), a highly hydrophobic transmembrane domain and a C-terminal
 intracellular domain that also shows homology to EGFR. Its
 overexpression correlates with a poor prognosis in breast and
 ovarian cancers. The invention provides Her-2/neu fusion
 proteins, nucleic acids encoding them, viral vectors, and vaccines

CC comprising the fusion proteins or nucleic acid molecules. In
 CC preferred fusion proteins, the extracellular domain of a Her-2/neu
 CC protein is fused to a Her-2/neu intracellular domain or
 CC phosphorylation domain (or its Deltapp fragment). An immune
 CC response to Her-2/neu protein is elicited or enhanced by
 CC administering the fusion protein in the form of a vaccine, or by
 CC transfecting cells of an animal ex vivo with a nucleic acid
 CC encoding the fusion protein, and delivering the transfected cells
 CC to the animal. The fusion proteins, nucleic acids, and isolated
 CC specific T-cells are useful for inhibiting the development of a
 CC cancer, especially breast, ovarian, colon, lung or prostate cancer
 CC in a patient. T cells that specifically react with a Her-2/neu
 CC fusion protein can be used to remove tumour cells from a sample in
 CC order to inhibit the development of cancer in a patient.

XX Sequence 1255 AA;

Query Match 98.7%; Score 6714; DB 23; Length 1255;
 Best Local Similarity 98.8%; Pred. No. 0;
 Matches 1240; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLLALLPPGAASSTVCTGTDMLRLPASPTHLMRLHYQGCQVVGNL 60
 Db 1 MELAALCRWGLLLALLPPGAASSTVCTGTDMLRLPASPTHLMRLHYQGCQVVGNL 60
 Qy 61 ELTYLPTNASLSFLQDIQEVQGVLIHNVQVPLQRLRIVRGTOLEFDNVALAVLDNG 120
 Db 61 ELTYLPTNASLSFLQDIQEVQGVLIHNVQVPLQRLRIVRGTOLEFDNVALAVLDNG 120
 Qy 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDITLWKDIFHKNNQLA 180
 Db 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDITLWKDIFHKNNQLA 180
 Qy 181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTYCAGGCARCKGPLPTDCHEQC 240
 Db 181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTYCAGGCARCKGPLPTDCHEQC 240
 Qy 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVQYIKANSKFGITELRTYFGASCVTACP 300
 Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVQYIKANSKFGITELRTYFGASCVTACP 300
 Qy 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTORCEKCKPCARVCYGLGMEHLREVRVTSAN 360
 Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTORCEKCKPCARVCYGLGMEHLREVRVTSAN 360
 Qy 361 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFETLEEITGVLYISAWPDSLP 420
 Db 361 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFETLEEITGVLYISAWPDSLP 420
 Qy 421 DLSVFONLQVIRGRILHNGAYSLLTQGLIGISWGLRSLRELGLGLAIHHNTHLCFVHTV 480
 Db 421 DLSVFONLQVIRGRILHNGAYSLLTQGLIGISWGLRSLRELGLGLAIHHNTHLCFVHTV 480
 Qy 481 PMDQLFRNPHQALLHTANRPEDECVCGEGLACHQLCARGHCWGPGPTQCVCNCSQFLRGQEC 540
 Db 481 PMDQLFRNPHQALLHTANRPEDECVCGEGLACHQLCARGHCWGPGPTQCVCNCSQFLRGQEC 540
 Qy 541 VEECRVLQGLPREYVNAHCLPCHPECQONGSVTCFGEADOCVCAHYKDPFFCVARC 600
 Db 541 VEECRVLQGLPREYVNAHCLPCHPECQONGSVTCFGEADOCVCAHYKDPFFCVARC 600
 Qy 601 PSGVKPDLGYMPIWPKFPDEEGACOPCPINCTHSCVDLDDKGPAPORASPLTSIVSAVVG 660
 Db 601 PSGVKPDLGYMPIWPKFPDEEGACOPCPINCTHSCVDLDDKGPAPORASPLTSIVSAVVG 660
 Qy 661 ILLVVVLGVVFGILIKRRQOKIRKYTMRELLQETELVEPLTPSGAMPNQAQMRILKETEL 720
 Db 661 ILLVVVLGVVFGILIKRRQOKIRKYTMRELLQETELVEPLTPSGAMPNQAQMRILKETEL 720
 Qy 721 RKVKVLGSGAFGTVYKGIWIPGENVKIPVAIKVLRNTSPKANKIILDEAYMAGVGP 780
 Db 721 RKVKVLGSGAFGTVYKGIWIPGENVKIPVAIKVLRNTSPKANKIILDEAYMAGVGP 780

QY 781 YVSRLLGICLTSTVQVLTQMPYCCLLDHYVRENRGRIGSODLLNWCMIKAGMSYLEDVR 840
Db 781 YVSRLLGICLTSTVQVLTQMPYCCLLDHYVRENRGRIGSODLLNWCMIKAGMSYLEDVR 840
QY 841 LVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGGKVPKIMMALESILRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGGKVPKIMMALESILRRFT 900
QY 901 HQSDVWSYGVTVWELMTFGAKYDGPAREIPDLLEKGERLPQPICTIDVYIMVWKWM 960
Db 901 HQSDVWSYGVTVWELMTFGAKYDGPAREIPDLLEKGERLPQPICTIDVYIMVWKWM 960
QY 961 IDSECRPRFELVSEFSEMRDQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMGDLVDA 1020
Db 961 IDSECRPRFELVSEFSEMRDQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMGDLVDA 1020
QY 1021 EYLVPQOQFFCDDPAPGAGWVHRRSSSTRSGGDLTLGLPSEEEAPRPLAPSEG 1080
Db 1021 EYLVPQOQFFCDDPAPGAGWVHRRSSSTRSGGDLTLGLPSEEEAPRPLAPSEG 1080
QY 1081 AGSDVDFDGLGMAAKGLQSLPTHDPSPLOQYSEDPTVPLPSETDGYVAPLTCSPPQEVV 1140
Db 1081 AGSDVDFDGLGMAAKGLQSLPTHDPSPLOQYSEDPTVPLPSETDGYVAPLTCSPPQEVV 1140
QY 1141 NOPDVRPPSPREGPLPAAPAGATLERAKTILSPGKNGVVKDVFAPGGAVENPEYLTPO 1200
Db 1141 NOPDVRPPSPREGPLPAAPAGATLERAKTILSPGKNGVVKDVFAPGGAVENPEYLTPO 1200
QY 1201 GGAAPQHPPPAPSPAFDNLVYWDQDPPERGAPPSTFKGTAEENPEYLGDLVVPV 1255
Db 1201 GGAAPQHPPPAPSPAFDNLVYWDQDPPERGAPPSTFKGTAEENPEYLGDLVVPV 1255

RESULT 14

AAU77114
ID AAU77114 standard; Protein; 1255 AA.
XX AAU77114;
XX
XX
DT 05-JUN-2002 (first entry)
XX Human Her-2/neu polypeptide.
DE
XX
KW Human; Her-2/neu; cytostatic; haematological malignancy; CML;
KW acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;
KW chronic lymphocytic leukaemia; mveloma; non-Hodgkin's lymphoma; MDS;
KW Hodgkin's lymphoma; T cell therapy.
XX
OS Homo sapiens.
XX
XX WO200213847-A2.
XX
XX 21-FEB-2002.
XX
XX 13-AUG-2001; 2001WO-US25408.
XX
XX 14-AUG-2000; 2000US-0638280.
XX 28-SEP-2000; 2000US-0675904.
XX
XX (CORI-) CORIXA CORP.
XX
XX Gaiger A, Cheever MA, Hand-zimmermann S;
XX WPI; 2002-280741/32.
XX N-PSDB; ABK10730.
XX
XX Inhibiting haematological malignancy development by administering
PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
PT encoding the polypeptide, or antigen presenting cells expressing the
PT polypeptide
XX
XX
PS Disclosure; Page 71-74; 74pp; English.
XX

CC The invention relates to a method for inhibiting development of
CC haematological malignancy in a patient by administering a polypeptide
CC comprising an immunogenic portion of Her-2/neu or a polynucleotide
CC encoding the polypeptide. Antigen presenting cells that express the
CC protein can also be administered. The sequences are used for inhibiting
CC development of haematological malignancy such as acute myelogenous
CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
CC lymphoma. This sequence represents the human Her-2/neu polypeptide.
XX
SQ Sequence 1255 AA;

Query Match 98.7%; Score 6714; DB 23; Length 1255;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 3; Mismatches 12; Indels 0; Gaps 0;
QY 1 MELAALCRWGLLALLPPGAASQVCTGDMKLRLPASPTHLDMLRHLYQGCVVQGNL 60
Db 1 MELAALCRWGLLALLPPGAASQVCTGDMKLRLPASPTHLDMLRHLYQGCVVQGNL 60
QY 61 ELTYLPTNASLSFLQDIQEVQGVLIHNVQVPLQRLRIVRGTLQDFDNVALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIQEVQGVLIHNVQVPLQRLRIVRGTLQDFDNVALAVLDNG 120
QY 121 DPLNNTPTVTGASPGSLRELQRLSLTEILKGGVLIQRPOLCYQDITLWKDIFHKNNQLA 180
Db 121 DPLNNTPTVTGASPGSLRELQRLSLTEILKGGVLIQRPOLCYQDITLWKDIFHKNNQLA 180
QY 181 LTLIDNRRACHPSCPMCKSGRCWGESSEDCQSLTRTVGAGGCARCKGPLTDCCHEQC 240
Db 181 LTLIDNRRACHPSCPMCKSGRCWGESSEDCQSLTRTVGAGGCARCKGPLTDCCHEQC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVQYIKANSKFITELRTYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVQYIKANSKFITELRTYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPHNLQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNYLSTDVGSCTLVCPHNLQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360
QY 361 IOEFAGCKKIFGSLAFIPESFDGDPASNTAPLOPQLOVTFEETLEETIGLYISAWPDSL 420
Db 361 IOEFAGCKKIFGSLAFIPESFDGDPASNTAPLOPQLOVTFEETLEETIGLYISAWPDSL 420
QY 421 DLSVFONLVIRGRILHNGAYSILTIQGLGISWGLRSRLRELGLAIHNNHLCFVHTV 480
Db 421 DLSVFONLVIRGRILHNGAYSILTIQGLGISWGLRSRLRELGLAIHNNHLCFVHTV 480
QY 481 PWDQLFRNPQALLHTANRPEDECVGEGGLACHOLCARGHCWGPPTQCVNCSQFLRGQEC 540
Db 481 PWDQLFRNPQALLHTANRPEDECVGEGGLACHOLCARGHCWGPPTQCVNCSQFLRGQEC 540
QY 541 VEECRVLQGLPREYVNAHCLPCHPECOPNGSVTCFGEADQCACAHYKDPFPFCVARC 600
Db 541 VEECRVLQGLPREYVNAHCLPCHPECOPNGSVTCFGEADQCACAHYKDPFPFCVARC 600
QY 601 PSGVKPDLISYMPIWKFPDEEGACQPCPNCTHSCVDLDDKQCPAEORASPLTSIVSAVVG 660
Db 601 PSGVKPDLISYMPIWKFPDEEGACQPCPNCTHSCVDLDDKQCPAEORASPLTSIVSAVVG 660
QY 661 ILLVVLGVVFGILIKRQOKIRKYMRLLOTELVEPLTFSGAMPNQAOMRIKETEEL 720
Db 661 ILLVVLGVVFGILIKRQOKIRKYMRLLOTELVEPLTFSGAMPNQAOMRIKETEEL 720
QY 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMAGVGP 780
Db 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMAGVGP 780
QY 781 YVSRLLGICLTSTVQVLTQMPYCCLLDHYVRENRGRIGSODLLNWCMIKAGMSYLEDVR 840
Db 781 YVSRLLGICLTSTVQVLTQMPYCCLLDHYVRENRGRIGSODLLNWCMIKAGMSYLEDVR 840
QY 841 LVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGGKVPKIMMALESILRRFT 900

Db 841 LVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEVHADGGKVPKIMWALESILRRFT 900
QY 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVWKWM 960
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVWKWM 960
QY 961 IDSECRPRFRELVSFMRMARDPQFVVIQNEIDLGPASPLDSTFYRSLLEDDDDMDGLVDA 1020
Db 961 IDSECRPRFRELVSFMRMARDPQFVVIQNEIDLGPASPLDSTFYRSLLEDDDDMDGLVDA 1020
QY 1021 EYLVPQGFPCPDPAAGAGWVHRRSSSTRSGGDLTLGLEPSEEEAPRPLAPSEG 1080
Db 1021 EYLVPQGFPCPDPAAGAGWVHRRSSSTRSGGDLTLGLEPSEEEAPRPLAPSEG 1080
QY 1081 AGSDVFDGDLGMAAKGLQSLPHDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAAKGLQSLPHDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
QY 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVVDVAFGGAVENPEYLTPO 1200
Db 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVVDVAFGGAVENPEYLTPO 1200
QY 1201 GGAAPQHPPPAFSPAFDNLVYWDQDPPPERGAPPSTFKGTPTAENPEYLGIDVPV 1255
Db 1201 GGAAPQHPPPAFSPAFDNLVYWDQDPPPERGAPPSTFKGTPTAENPEYLGIDVPV 1255

RESULT 15

AAR39568
ID AAR39568 standard; Protein; 1433 AA.
XX
AC AAR39568;
XX
DT 07-FEB-1994 (first entry)
XX
DE Sequence of c-erbB-2 tumour antigen.
XX
KW Tumour antigen; c-erbB-2; glycoprotein.
XX
OS Homo sapiens.
XX
PN WO9316185-A.
XX
PD 19-AUG-1993.
XX
PF 05-FEB-1993; 93WO-US01055.
XX
PR 06-FEB-1992; 92US-0831967.
XX
PA (CETU) CETUS ONCOLOGY CORP.
PA (CREA-) CREATIVE BIOMOLECULES INC.
XX
PI Houston LL, Huston JS, Oppermann H, Ring DB;
XX
DR WPI; 1993-272889/34.
DR N-PSDB; AAR39568.

PT New single chain Fv polypeptide binding to C-erbB-2 tumour
PT antigen - for imaging or treating breast or ovarian cancer etc.
XX
PS Disclosure; pages 48-54; 87pp; English.
XX
CC c-erbB-2 refers to a protein antigen expressed on the surface of
CC tumour cells, such as breast and ovarian tumour cells, which is an
CC approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric
CC pt. of about 5.3 (see AAR39568). X in AAR39568 represents
CC the location of a stop codon in AAR39568.
XX
SQ Sequence 1433 AA;

Query Match 98.1%; Score 6671; DB 14; Length 1433;
Best Local Similarity 98.2%; Pred. No. 0;

Matches 1233; Conservative 5; Mismatches 17; Indels 0; Gaps 0;
QY 1 MELAALCRGLGLALLAPPGAASQVCTGDMKLRLPASPTHLDMLRHLYQGCQVQGNL 60
Db 1 MELAALCRGLGLALLAPPGAASQVCTGDMKLRLPASPTHLDMLRHLYQGCQVQGNL 60
QY 61 ELYLPTNASLSFLQDIQEVQGVYLLAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
Db 61 ELYLPTNASLSFLQDIQEVQGVYLLAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
QY 121 DPLNNTPTVGTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTRILKWDIFHKKNOLA 180
Db 121 DPLNNTPTVGTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTRILKWDIFHKKNOLA 180
QY 181 LTLIDNTRRACHPCSPMCKGRCWGESSEDCOSLTRTVCAGGCARCKGPLPTDCCHEQC 240
Db 181 LTLIDNTRRACHPCSPMCKGRCWGESSEDCOSLTRTVCAGGCARCKGPLPTDCCHEQC 240
QY 241 AAGCTGPKHSDCLACILHFNHSGICELHCPALVQYIKANSKFIGITELRTYFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACILHFNHSGICELHCPALVQYIKANSKFIGITELRTYFGASCVTACP 300
QY 301 YNYLSTDVSGSCTLVCPHNOEVTAEQTCRCKSKPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNYLSTDVSGSCTLVCPHNOEVTAEQTCRCKSKPCARVCYGLGMEHLREVRVTSAN 360
QY 361 IOEFAGCKTIFGSLAEFLPESFDGDPASNTAPLOPELQVFEETLEETGVLYISAWPDSL 420
Db 361 IOEFAGCKTIFGSLAEFLPESFDGDPASNTAPLOPELQVFEETLEETGVLYISAWPDSL 420
QY 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGISWGLRSLRELGSGLALIHNNTHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGISWGLRSLRELGSGLALIHNNTHLCFVHTV 480
QY 481 PMDQLFRNPHQALLHTANRPEDECVGEGLACHOLCARGHCWGPGTQCVNCSOFLRGQEC 540
Db 481 PMDQLFRNPHQALLHTANRPEDECVGEGLACHOLCARGHCWGPGTQCVNCSOFLRGQEC 540
QY 541 VEECRVLQGLPREYNARHCLPCHPSCQPNQSVTCFGEADOCVACAHKDPFVCVARC 600
Db 541 VEECRVLQGLPREYNARHCLPCHPSCQPNQSVTCFGEADOCVACAHKDPFVCVARC 600
QY 601 PSQVKPDLSPMPTWPKFPPDEGACQPCPINCTHSCVDLDDKGPAPQASPLTSISAVVG 660
Db 601 PSQVKPDLSPMPTWPKFPPDEGACQPCPINCTHSCVDLDDKGPAPQASPLTSISAVVG 660
QY 661 ILLVVVLGVVFGILIKRQOKIRKYTMRLLOSTELVEPLTPSGAMPNOAQMRLKETEL 720
Db 661 ILLVVVLGVVFGILIKRQOKIRKYTMRLLOSTELVEPLTPSGAMPNOAQMRLKETEL 720
QY 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKILDEAYVMAGVSP 780
Db 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKILDEAYVMAGVSP 780
QY 781 YVSRLLGICLTSTVQLVTQMLPYCCLLDHVRNRRGLSGQDLNWCWQIAKNGSYLEDVR 840
Db 781 YVSRLLGICLTSTVQLVTQMLPYCCLLDHVRNRRGLSGQDLNWCWQIAKNGSYLEDVR 840
QY 841 LVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEVHADGGKVPKIMWALESILRRFT 900
Db 841 LVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEVHADGGKVPKIMWALESILRRFT 900
QY 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVWKWM 960
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVWKWM 960
QY 961 IDSECRPRFRELVSFMRMARDPQFVVIQNEIDLGPASPLDSTFYRSLLEDDDDMDGLVDA 1020
Db 961 IDSECRPRFRELVSFMRMARDPQFVVIQNEIDLGPASPLDSTFYRSLLEDDDDMDGLVDA 1020
QY 1021 EYLVPQGFPCPDPAAGAGWVHRRSSSTRSGGDLTLGLEPSEEEAPRPLAPSEG 1080
Db 1021 EYLVPQGFPCPDPAAGAGWVHRRSSSTRSGGDLTLGLEPSEEEAPRPLAPSEG 1080

```
Qy 1081 AGSDVFDGDLGMGAAGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
    |||||
Db 1081 AGSDVFDGDLGMGAAGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
    |||||
Qy 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKOVFAFGGAVENPEYLTPO 1200
    |||||
Db 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKOVFAFGGAVENPEYLTPO 1200
    |||||
Qy 1201 GGAAPQHPHPPAFSPAFDNLYYWDQDPPPERGAPSTFKGTPTAENPEYLGLDVVPV 1255
    |||||
Db 1201 GGAAPQHPHPPAFSPAFDNLYYWDQDPPPERGAPSTFKGTPTAENPEYLGLDVVPV 1255
    |||||
```

Search completed: July 22, 2003, 08:41:01
Job time : 43.9774 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:11:38 ; Search time 37.9774 Seconds
(without alignments)
4403.399 Million cell updates/sec

Title: SEQ4-59-73-12

Perfect score: 6809

Sequence: 1 MELALCRWGLLLALLPFGA.....TFKGTPTAENPEYLGDLVPV 1255

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq.101002.*

1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6735	98.9	1255	21	AAE240620 Human heregulin 2
2	6735	98.9	1255	22	AAE12130 Human tyrosine kin
3	6735	98.9	1255	22	AAE60167 HER2 transgene pla
4	6735	98.8	1255	23	AAU74545 Human HER2 (ErbB2)
5	6729	98.8	1255	17	AAW01111 HER-2/neu protein.
6	6729	98.8	1255	20	AAW92406 Human HER-2/neu on
7	6729	98.8	1255	21	AAE21198 Human HER-2/neu pr
8	6729	98.8	1255	21	AAE4780 Amino acid sequenc
9	6729	98.8	1255	22	AAE85458 Human HER-2/neu pr
10	6729	98.8	1255	22	AAE88267 HER2/neu amino aci

11	6729	98.8	1255	23	AAE24067 Human Her-2 protei
12	6729	98.8	1255	23	AAE20479 Human Her-2/neu pr
13	6729	98.8	1255	23	AAE51143 Human Her-2/neu on
14	6729	98.8	1255	23	AAU77114 Human Her-2/neu po
15	6686	98.2	1433	14	AAE39568 Sequence of c-erbB
16	6565	96.4	1223	23	AAU98923 Human breast cance
17	6412	94.2	1200	21	AAE21208 Human HER-2/neu pr
18	5929.5	87.1	1256	21	AAE21199 Rat HER-2/neu prot
19	5929.5	87.1	1256	23	AAE51144 Rat Her-2/neu onco
20	5909.5	86.8	1256	21	AAE21206 Mouse Her-2/neu pr
21	5909.5	86.8	1256	22	AAE62860 Amino acid sequenc
22	5909.5	86.8	1256	23	AAE51151 Mouse Her-2/neu on
23	4815	70.7	919	21	AAE21203 Human HER-2/neu fu
24	4815	70.7	919	23	AAE51148 Her-2/neu extracel
25	4059.5	59.6	920	23	AAE51152 Mouse Her-2/neu ex
26	4059.5	59.6	926	23	AAE51153 Mouse Her-2/neu ex
27	3699	54.3	712	21	AAE21204 Human HER-2/neu fu
28	3699	54.3	712	23	AAE51149 Her-2/neu extracel
29	3553	52.2	782	18	AAE19764 Her2-GM-CSF immuno
30	3551	52.2	653	21	AAE21200 Extracellular HER-
31	3551	52.2	653	23	AAE51145 Human Her-2/neu on
32	3513	51.6	645	22	AAE60408 Human ErbB2 oncopr
33	3513	51.6	645	22	AAE61593 Human ErbB2 extrac
34	3448	50.6	951	21	AAE44993 DC8cFv-erbB2EC fu
35	3345	49.1	624	11	AAE08222 Extracellular port
36	3117	45.8	1210	21	AAE19259 Amino acid sequenc
37	3117	45.8	1210	21	AAE50616 Human EGF receptor
38	3117	45.8	1210	23	AAE23019 Human Her-1 protei
39	3117	45.8	1210	23	AAE50768 Human epidermal gr
40	3115	45.7	1210	22	AAE68420 Amino acid sequenc
41	3084	45.3	583	23	AAE20483 Human protein for
42	3084	45.3	587	23	AAE20481 Human protein for
43	3083	45.3	599	23	AAE20484 Human protein for
44	3083	45.3	600	23	AAE20482 Human protein for
45	3076	45.2	1210	23	ABE51768 Human epidermal gr

ALIGNMENTS

RESULT 1

AAE240620
ID AAY92620 standard; Protein; 1255 AA.
XX
AC AAY92620;
XX
DT 10-AUG-2000 (first entry)
XX
DE Human heregulin 2 (Her2).
XX
KW Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;
KW self-protein; cancer; breast cancer; prostate cancer;
KW cell-associated peptide antigen; foreign epitope.
XX
OS Homo sapiens.
XX
PH Key
FT Domain
FT
FT Location/Qualifiers
FT /label= N-terminal
FT /note= "mature polypeptide"
FT 5...25
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 59...73
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 103...117
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 149...163
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 174...323
FT Domain

FT		/label= Cysteine_rich_domain
FT	Region	210..224
FT		/label= insertion region
FT		/note= "suitable for foreign epitope insertion"
FT	Region	250..264
FT		/label= insertion region
FT		/note= "suitable for foreign epitope insertion"
FT	Domain	324..483
FT		/label= Ligand_binding_domain
FT	Region	325..339
FT		/label= insertion region
FT		/note= "suitable for foreign epitope insertion"
FT	Region	369..383
FT		/label= insertion region
FT		/note= "suitable for foreign epitope insertion"
FT	Region	465..479
FT		/label= insertion region
FT		/note= "suitable for foreign epitope insertion"
FT	Domain	484..623
FT		/label= Cysteine_rich_domain
FT	Region	579..593
FT		/label= insertion region
FT		/note= "suitable for foreign epitope insertion"
FT	Domain	624..654
FT		/label= Transmembrane_domain
FT	Region	632..652
FT		/label= insertion region
FT		/note= "suitable for foreign epitope insertion"
FT	Region	653..667
FT		/label= insertion region
FT		/note= "suitable for foreign epitope insertion"
FT	Domain	655..1010
FT		/label= Tyrosine_kinase_domain
FT	Region	661..675
FT		/label= insertion region
FT		/note= "suitable for foreign epitope insertion"
FT	Region	695..709
FT		/label= insertion region
FT		/note= "suitable for foreign epitope insertion"
FT	Region	710..730
FT		/label= insertion region
FT		/note= "suitable for foreign epitope insertion"
FT	Domain	1011..1235
FT		/label= C-terminal_domain
XX		
XX	W0200020027-A2.	
XX		
XX	13-APR-2000.	
XX		
XX	05-OCT-1999;	99WO-DK00525.
XX		
PR	05-OCT-1998;	98DK-0001261.
PR	20-OCT-1998;	98US-0105011.
XX		
PA	(MEBI-) M & E BIOTECH AS.	
XX		
PI	Steinaa L, Mouriteen S, Nielsen KG, Haaning J, Leach D, Dalum I;	
PI	Gautam A, Birk P, Karlsson G;	
XX		
XX	WPI; 2000-349917/30.	
DR	N-PSDB; AAA09455.	
XX		
PT	Inducing immune responses to weakly immunogenic, tumor associated	
PT	peptide antigens for the treatment of breast and prostate cancer	
XX		
XX	Claim 62; Page 193-198; 220pp; English.	
XX		
CC	This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of	
CC	Her2 can be used in the claimed method as an autovaccine to induce a CTL	
CC	response. Subdominant CRL epitopes, antibody binding regions and	
CC	cysteine residues involved in disulfide bonds are preserved in the	
CC	immunogenized forms. Regions suitable for the insertion of foreign T	
CC	helper epitopes were identified (see features table). The method	

CC	is used for inducing immune responses against weakly immunogenic
CC	cell-associated peptide antigens (PA) such as those associated with
CC	cancers (self-proteins), e.g. human prostate specific membrane antigen
CC	(PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).
CC	The method comprises effecting simultaneous presentation by antigen
CC	producing cells (APCs) of the animals immune system of: (1) at least 1
CC	CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1
CC	B-cell group derived from the cell-associated PA; and (2) at least 1
CC	first T helper cell group which is foreign to the animal. Analogues of
CC	human PSM, human Her2 and human/murine FGF8b comprising a substantial
CC	part of all known and predicted CTL and B-cell epitopes of the respective
CC	PA and including at least one foreign T helper epitope are also claimed.
CC	The method is used to treat prostate, prostate/breast or breast cancer.
CC	when the PA is human PSM, FGF8b and Her2, respectively.
xx	
SQ	Sequence 1255 AA;
Query Match	98.9%; Score 6735; DB 21; Length 1255;
Best Local Similarity	99.0%; Pred. No. 0;
Matches 1242; Conservative	2; Mismatches 11; Indels 0; Gaps 0;
Qy	1 MELAALCRWGLLLALLPPGAASQVCTGTDMLKRLPASPETHLDMLRLHYQGQCVVQGNL 60
Db	1 MELAALCRWGLLLALLPPGAASQVCTGTDMLKRLPASPETHLDMLRLHYQGQCVVQGNL 60
Qy	61 ELYVLPNASLSFLQDIQEVOQYIKANSKFIGITELQRLRIVRGQTLPEDNYALAVLDNG 120
Db	61 ELYVLPNASLSFLQDIQEVOGVYLAHNNQVRVQLQRLRIVRGQTLPEDNYALAVLDNG 120
Qy	121 DPLNNTPTVTGASGGREQLQRLSLTEILAGGVLIQRNPOLCYQDITLWKDIFHKNNQLA 180
Db	121 DPLNNTPTVTGASGGREQLQRLSLTEILAGGVLIQRNPOLCYQDITLWKDIFHKNNQLA 180
Qy	181 LTIDITNRSRACHPCSPMKGRCGWSSESDCQSLTRTVCGGCARCKGPLTDCCHQC 240
Db	181 LTIDITNRSRACHPCSPMKGRCGWSSESDCQSLTRTVCGGCARCKGPLTDCCHQC 240
Qy	241 AAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDTTFESMPNPEGRTYFGASCVTACP 300
Db	241 AAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDTTFESMPNPEGRTYFGASCVTACP 300
Qy	301 YNYLSTDVGCTLVCPLHNOEVTAEDGTQCEKSKPCARVCYGLGMEHLREVRVTSAN 360
Db	301 YNYLSTDVGCTLVCPLHNOEVTAEDGTQCEKSKPCARVCYGLGMEHLREVRVTSAN 360
Qy	361 IQEPAGCKKIFGSLAFIPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSL 420
Db	361 IQEPAGCKKIFGSLAFIPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSL 420
Qy	421 DLSVFQNLQVIRGRILHNGAYSILTQGLGTSWLGSLRLSGSLALIHNNTHLCFVHTV 480
Db	421 DLSVFQNLQVIRGRILHNGAYSILTQGLGTSWLGSLRLSGSLALIHNNTHLCFVHTV 480
Qy	481 PWDQLFRNPHQALLHTANRDEDECVGSGIACHOLCARGHCWGPGPTQCVNCSOFLRGQEC 540
Db	481 PWDQLFRNPHQALLHTANRDEDECVGSGIACHOLCARGHCWGPGPTQCVNCSOFLRGQEC 540
Qy	541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVCAHYKDPFPCVARC 600
Db	541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVCAHYKDPFPCVARC 600
Qy	601 PSGVKPDLSYMPIWKFPDEGACOPCPINCTHSCVDLDDKCPAEQBPASPLTSTVSAVG 660
Db	601 PSGVKPDLSYMPIWKFPDEGACOPCPINCTHSCVDLDDKCPAEQBPASPLTSTVSAVG 660
Qy	661 ILLVVVLGVVFGILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMNPQAMRLIKETEL 720
Db	661 ILLVVVLGVVFGILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMNPQAMRLIKETEL 720
Qy	721 RKVKVLGSGAFVTYKGIWIPDGENVKIPVAIKVLRRENTSPKANKEILDEAYVMAVGSP 780
Db	721 RKVKVLGSGAFVTYKGIWIPDGENVKIPVAIKVLRRENTSPKANKEILDEAYVMAVGSP 780

QY 781 YVSRLLGICLTSTVLTQMPYGCCLLDHVRNRRGLSQDILLNWCMIKAGMSYLEVDVR 840
 Db 781 YVSRLLGICLTSTVLTQMPYGCCLLDHVRNRRGLSQDILLNWCMIKAGMSYLEVDVR 840
 QY 841 LVHRDLAARNVLKSPNHNKVTDFGLARLLDDIDEYHADGKVPKIMWMALESILRRFT 900
 Db 841 LVHRDLAARNVLKSPNHNKVTDFGLARLLDDIDEYHADGKVPKIMWMALESILRRFT 900
 QY 901 HQSDWMSYGVTVWELMTGAKPYDGPAREIPDLLEKGERLPQPPICTIDVYIMVKWM 960
 Db 901 HQSDWMSYGVTVWELMTGAKPYDGPAREIPDLLEKGERLPQPPICTIDVYIMVKWM 960
 QY 961 IDSECRPRFRLVSEFSRMARDPQRFVVIQNEIDGPASPLSTFYRSLLEDMDGLVDA 1020
 Db 961 IDSECRPRFRLVSEFSRMARDPQRFVVIQNEIDGPASPLSTFYRSLLEDMDGLVDA 1020
 QY 1021 EYLVPQOGFFCPDPAFCAGGMVHRHRSSTRSGGDLTLGLPSEBEPAPSLAPSEG 1080
 Db 1021 EYLVPQOGFFCPDPAFCAGGMVHRHRSSTRSGGDLTLGLPSEBEPAPSLAPSEG 1080
 QY 1081 AGSDVFDGDLGWAAGKLOSLTPHDPSPLORYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
 Db 1081 AGSDVFDGDLGWAAGKLOSLTPHDPSPLORYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
 QY 1141 NOPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVDVPAFCGAVENPEYLTPO 1200
 Db 1141 NOPDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVDVPAFCGAVENPEYLTPO 1200
 QY 1201 GGAAPQHPHPPAFSPADNLVYWDQDPPERCAPPSTPKGTPTAENPEYGLDVPV 1255
 Db 1201 GGAAPQHPHPPAFSPADNLVYWDQDPPERCAPPSTPKGTPTAENPEYGLDVPV 1255

RESULT 2

AAE12130
 ID AAE12130 standard; Protein; 1255 AA.

XX AC AAE12130;

XX DT 18-DEC-2001 (first entry)

XX DE Human tyrosine kinase-type receptor, HER-2.

XX KW Therapeutic compound; major histocompatibility complex; vaccine;
 KW antigenic peptide; MHC; immunoregulatory; immune response; HER-2;
 KW adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;
 KW antigen presenting cell; human; tyrosine kinase-type receptor.

XX OS Homo/sapiens.

XX FH Key Location/Qualifiers
 XX Region 774..782
 XX FT /note= "Antigenic epitope"

XX FN WO200168677-A2.

XX PD 20-SEP-2001.

XX PF 16-MAR-2001; 2001WO-US40328.

XX PR 16-MAR-2000; 2000US-0527487.

XX XX (GENZ) GENZYME CORP.

XX PA Nicolette CA;

XX PI Nicolette CA;

XX XX WPI; 2001-616284/71.

XX DR N-PSDB; AAD19731.

XX XX Novel synthetic therapeutic compound for inducing immune response and
 PT for use in adoptive immunotherapy, has enhanced binding to major
 PT histocompatibility molecules and enhanced immunoregulatory properties

PT

XX

PS Claim 4; Page 63-67; 69pp; English.

XX The invention relates to synthetic therapeutic compounds (antigenic
 CC peptides) with enhanced binding to major histocompatibility complex
 CC (MHC) molecules and enhanced immunoregulatory properties relative
 CC to their natural counterparts. Compounds of the invention are useful
 CC for inducing an immune response in a subject and for use in adoptive
 CC immunotherapy. They are useful as components of anti-cancer vaccines
 CC and to expand immune effector cells that are specific for cancers
 CC characterised by expression of the breast cancer antigen, HER-2.
 CC Polynucleotides that encode peptides of the invention are useful as
 CC hybridisation probes and as primers for the detection of genes of gene
 CC transcripts that are expressed in antigen presenting cells (APCs), to
 CC confirm transduction of polynucleotides into host cells. The present
 CC sequence is human tyrosine kinase-type receptor, HER-2. Compounds
 CC of the invention are designed based on the HER-2 antigenic peptide
 CC (774-782).

XX SQ Sequence 1255 AA;

Query Match 98.9%; Score 6735; DB 22; Length 1255;
 Best Local Similarity 99.0%; Pred. No. 0;
 Matches 1242; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLALLPPGAASQVCTGTDMLRLPASPEHDLMLRHLVGGCQVQGNL 60

Db 1 MELAALCRWGLLALLPPGAASQVCTGTDMLRLPASPEHDLMLRHLVGGCQVQGNL 60

QY 61 ELYLPNTNASLSFLQDIOEQVQYIKANSKFIGITELQRLIRIVRGTLFEDNYALAVLDNG 120

Db 61 ELYLPNTNASLSFLQDIOEQVQYIAHNOVPLQRLIRIVRGTLFEDNYALAVLDNG 120

QY 121 DPLNNTTPVTGASPGGLRELQLESLEILKGGVLIORNPOLCYQDTTLWKDI FHKNNOLA 180

Db 121 DPLNNTTPVTGASPGGLRELQLESLEILKGGVLIORNPOLCYQDTTLWKDI FHKNNOLA 180

QY 181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVCCAGCARCKGLPTDCCHEQC 240

Db 181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVCCAGCARCKGLPTDCCHEQC 240

QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMNPPEGRYTFGASCVTACP 300

Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMNPPEGRYTFGASCVTACP 300

QY 301 YNYLSTDVGSCTLVCPLHNOEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360

Db 301 YNYLSTDVGSCTLVCPLHNOEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360

QY 361 IQEFAGCKKIFGSLAFPLPESFDGDPASNTAPLOPEQLQVFPETLEEITGYLYISAWPDSL 420

Db 361 IQEFAGCKKIFGSLAFPLPESFDGDPASNTAPLOPEQLQVFPETLEEITGYLYISAWPDSL 420

QY 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGISWLGSLRSLRELGSGLALIHNTLFCVHTV 480

Db 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGISWLGSLRSLRELGSGLALIHNTLFCVHTV 480

QY 481 PWDOLFERNPHQALLHTANRPEDECVGEGGLACHOLCARGHCWGPGTQCVCNCSQFLRGQEC 540

Db 481 PWDOLFERNPHQALLHTANRPEDECVGEGGLACHOLCARGHCWGPGTQCVCNCSQFLRGQEC 540

QY 541 VEECRVLQGLFREYVVARHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFPFCVARC 600

Db 541 VEECRVLQGLFREYVVARHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFPFCVARC 600

QY 601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKCGCPAEQASPLTSIVSAVVG 660

Db 601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKCGCPAEQASPLTSIVSAVVG 660

QY 661 ILLVVLGVVFGILLIKRQOKIRKYTMRLLOSTELVEPLTTPSGAMPNQAQMRILKETEL 720

Db 661 ILLVVLGVVFGILLIKRQOKIRKYTMRLLOSTELVEPLTTPSGAMPNQAQMRILKETEL 720

```

Qy 721 RKVKVLSGARGTVYKGIWIPDGENVKIPVAIKVRENTSPKANKIILDEAYNVAGVSP 780
Db 721 RKVKVLSGARGTVYKGIWIPDGENVKIPVAIKVRENTSPKANKIILDEAYNVAGVSP 780
Qy 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRGRIGSQDLLNMCQIAKMSYLEDDR 840
Db 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRGRIGSQDLLNMCQIAKMSYLEDDR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGKVPKIKWMALESILRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGKVPKIKWMALESILRRFT 900
Qy 901 HQSDVWSYGVTVWELMTFGAPYDIPAREIPDLLEKGERLPQPPICITIDVYMWKWM 960
Db 901 HQSDVWSYGVTVWELMTFGAPYDIPAREIPDLLEKGERLPQPPICITIDVYMWKWM 960
Qy 961 IDSECRPRFRELVSFSESMARDPQRFVVIQNEIDLGASPLDSTFYRSILLEDDMGDLVDA 1020
Db 961 IDSECRPRFRELVSFSESMARDPQRFVVIQNEIDLGASPLDSTFYRSILLEDDMGDLVDA 1020
Qy 1021 BEYLVPQGFCTPDPAPGAGWVHRRSSSTRSGGDLTLGLEPSEBEEAPRSLAPSEG 1080
Db 1021 BEYLVPQGFCTPDPAPGAGWVHRRSSSTRSGGDLTLGLEPSEBEEAPRSLAPSEG 1080
Qy 1081 AGSDVFDGLGWAAGLQSLPTHDPSPLORYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGLGWAAGLQSLPTHDPSPLORYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140
Qy 1141 NQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTQP 1200
Db 1141 NQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTQP 1200
Qy 1201 GGAAPQHPHPPAFSPFNLYWDDPPERGAPSTFKGTPTAENPEYGLDVPV 1255
Db 1201 GGAAPQHPHPPAFSPFNLYWDDPPERGAPSTFKGTPTAENPEYGLDVPV 1255

```

RESULT 3

AAB60167 standard; Protein; 1255 AA.

XX AC AAB60167;

XX DT 03-APR-2001 (first entry)

XX DE HER2 transgene plasmid construct encoded protein.

XX KW Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer; antibody.

XX OS Homo sapiens.

XX OS Synthetic.

XX PN WO200100244-A2.

XX PD 04-JAN-2001.

XX PF 23-JUN-2000; 2000WO-US17229.

XX PR 25-JUN-1999; 99US-0141316.

XX PR 16-MAR-2000; 2000US-0189844.

XX PA (GETH) GENENTECH INC.

XX XX Erickson S, Schwall R;

XX XX WPI; 2001-061962/07.

XX DR N-PSDB; AAF24297.

XX PT Treating tumors, particularly breast cancers, which overexpress an ErbB receptor and does not respond to an anti-ErbB antibody, comprises

XX PT conjugating the antibody to a maytansinoid -

PS Example 3; Fig 4; 92pp; English.

XX CC The present invention provides a method of treating cancer by administering a conjugate of anti-ErbB antibody with a maytansinoid. In particular, the antibody is directed against ErbB2 (also known as HER2 and p185neu). The method is particularly useful in the treatment of breast, ovarian, stomach, endometrial, salivary gland, lung, kidney, colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.

XX SQ Sequence 1255 AA;

Query Match 98.9%; Score 6735; DB 22; Length 1255;

Best Local Similarity 99.0%; Pred. No. 0;

Matches 1242; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLALALPPGAASQVCTGDMKRLPASPTHLDMLHLHYGCGQVQGNL 60

Db 1 MELAALCRWGLLALALPPGAASQVCTGDMKRLPASPTHLDMLHLHYGCGQVQGNL 60

Qy 61 ELYLPTNASLSFLQDIQEVQYIKANSKFITELQRLRVRGTQLFEDNYALAVLDNG 120

Db 61 ELYLPTNASLSFLQDIQEVQYILAHNVQVPLQRLRVRGTQLFEDNYALAVLDNG 120

Qy 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDI FHKNNQLA 180

Db 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDI FHKNNQLA 180

Qy 181 LTIIDNRSRACHPCSPMKSGRCWGESSEDCSLTRTVCGAGCARCKGLPTDCCHEQC 240

Db 181 LTIIDNRSRACHPCSPMKSGRCWGESSEDCSLTRTVCGAGCARCKGLPTDCCHEQC 240

Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTNTDTFESMPNPEGRYTFGASCVTACP 300

Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTNTDTFESMPNPEGRYTFGASCVTACP 300

Qy 301 YNYLSTDVSGCTLVCPLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRVTSAN 360

Db 301 YNYLSTDVSGCTLVCPLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRVTSAN 360

Qy 361 IQBFAGCKKIFGSLAFIPESFDGDPASNTAPLQEQLOVFTLEETIYGLIYISAWPDSLP 420

Db 361 IQBFAGCKKIFGSLAFIPESFDGDPASNTAPLQEQLOVFTLEETIYGLIYISAWPDSLP 420

Qy 421 DLSVFQNLQVIRGILHNGAYSLSLTQGLGISWLGRLSRLGSLALIHNTLHLCFVHTV 480

Db 421 DLSVFQNLQVIRGILHNGAYSLSLTQGLGISWLGRLSRLGSLALIHNTLHLCFVHTV 480

Qy 481 PWDOLFENPHQALLHTANRDECEVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Db 481 PWDOLFENPHQALLHTANRDECEVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540

Qy 541 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFVCARC 600

Db 541 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFVCARC 600

Qy 601 PSGVKPDLSYMPTWKFPDEEGACOPCPINCTHSCVDLDDKGCPEAQEASPLTSIVSAVVG 660

Db 601 PSGVKPDLSYMPTWKFPDEEGACOPCPINCTHSCVDLDDKGCPEAQEASPLTSIVSAVVG 660

Qy 661 ILLVVLGVVFGIILIKRRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAMRILKETEL 720

Db 661 ILLVVLGVVFGIILIKRRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAMRILKETEL 720

Qy 721 RKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVRENTSPKANKIILDEAYNVAGVSP 780

Db 721 RKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVRENTSPKANKIILDEAYNVAGVSP 780

Qy 781 YVSELLGICLTSTVQLVTQMLPYGCLLDHVRENRGRIGSQDLLNMCQIAKMSYLEDDR 840

Db 781 YVSELLGICLTSTVQLVTQMLPYGCLLDHVRENRGRIGSQDLLNMCQIAKMSYLEDDR 840

Qy 841 LVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGKVPKIKWMALESILRRFT 900

Db 841 LVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGKVPKIKWMALESILRRFT 900

Db 841 LVHRDLARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPDKWMALESILRRPT 900
 Qy 901 HQSDVMSYGVTVWELMTGAKPYDIPAREIPDLLEKGERLPQPPICTIDVYIMVKWM 960
 Db 901 HQSDVMSYGVTVWELMTGAKPYDIPAREIPDLLEKGERLPQPPICTIDVYIMVKWM 960
 Qy 961 IDSECRPRFRELVSFMSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLDDMGDLVDA 1020
 Db 961 IDSECRPRFRELVSFMSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLDDMGDLVDA 1020
 Qy 1021 EYLVPQGGFFCPDPAAGAGGWHHRHRSSTRSGGDLTLGLPSEBEAPRSPAPSEG 1080
 Db 1021 EYLVPQGGFFCPDPAAGAGGWHHRHRSSTRSGGDLTLGLPSEBEAPRSPAPSEG 1080
 Qy 1081 AGSDVFDGDLGMAAGKGLQSLTPHDSPLRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
 Db 1081 AGSDVFDGDLGMAAGKGLQSLTPHDSPLRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
 Qy 1141 NOPDVRPQPPREGPLPAARPAATLERAKTLSPGKNGVVDVPAGGAVENPEYLTPO 1200
 Db 1141 NOPDVRPQPPREGPLPAARPAATLERAKTLSPGKNGVVDVPAGGAVENPEYLTPO 1200
 Qy 1201 GGAAPQHPHPPAFSPADNLVYWDQDPPRGAPPSTFKGTPTAENPEYLGDLVDPV 1255
 Db 1201 GGAAPQHPHPPAFSPADNLVYWDQDPPRGAPPSTFKGTPTAENPEYLGDLVDPV 1255

RESULT 4

AAU74545
 ID AAU74545 standard; Protein; 1255 AA.
 XX
 AC AAU74545;
 XX
 DT 23-APR-2002 (first entry)
 XX
 DE Human HER2 (Erbb2) polypeptide.

XX Human; HER2; ErbB; epidermal growth factor receptor; receptor;
 KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
 KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
 KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
 KW glial disorder; astrocytal disorder; hypothalamic disorder;
 KW glandular disorder; macrophagal disorder; epithelial disorder;
 KW stromal disorder; blastocoealic disorder; inflammatory disorder;
 KW angiogenic disorder; immunological disorder.

XX Homo sapiens.

XX US2002001587-A1.

XX 03-JAN-2002.

XX 16-MAR-2001; 2001US-0811123.

XX 16-MAR-2000; 2000US-189844P.

PR 05-OCT-2000; 2000US-238327P.

XX (ERIC/) ERICKSON S.

PA (SCHW/) SCHWALL R.

PA (SLIW/) SLIKOWSKI M.

XX Erickson S, Schwall R, Slikowski M;

XX WPI; 2002-163686/21.

DR N-PSDB; ABK14058.

XX Treating tumour characterised by overexpression of epidermal growth
 FT factor receptor, ErbB or cancer in mammal, comprises administering
 PT anti-ErbB antibody-maytansinoid conjugate to the mammal -

XX Example 3; Fig 7; 93pp; English.

PS The invention relates to treating a tumour in a mammal, where the tumour

XX

CC is characterised by the overexpression of an epidermal growth factor
 CC receptor (ErbB) and does not respond or responds poorly, to treatment
 CC with an anti-ErbB antibody, comprising administering to the mammal an
 CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for
 CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
 CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
 CC prostate and bladder, preferably breast cancer. The breast cancer is a
 CC metastatic breast cancer or an aggressive form of metastatic breast
 CC cancer which overexpresses ErbB2. The method is also useful for treating
 CC neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal,
 CC epithelial, stromal, blastocoealic, inflammatory, angiogenic and
 CC immunological disorders. This sequence represents the human HER2 (Erbb2)
 CC polypeptide of the invention.

XX SQ Sequence 1255 AA;

Query Match 98.98; Score 6735; DB 23; Length 1255;
 Best Local Similarity 99.08; Pred. No. 0;
 Matches 1242; Conservative 2; Mismatches 11; Indels 0; Gaps 0;
 Qy 1 MELAALCRWGLLALLPPGAASQTCTGDMKRLPASPETHDMLRHLYGQCQVQGNL 60
 Db 1 MELAALCRWGLLALLPPGAASQTCTGDMKRLPASPETHDMLRHLYGQCQVQGNL 60
 Qy 61 ELYVLPNASLSFLQDIQEVQVQYIKANSKFIGITELQRLRIVRGTLQFEDNYALAVLNG 120
 Db 61 ELYVLPNASLSFLQDIQEVQVQYIKANSKFIGITELQRLRIVRGTLQFEDNYALAVLNG 120
 Qy 121 DPLNNTTPTVGTASPGGLRELQLRLSLTEILKGGVLIQRPOLCYQDTTLWKDIFHNKOLA 180
 Db 121 DPLNNTTPTVGTASPGGLRELQLRLSLTEILKGGVLIQRPOLCYQDTTLWKDIFHNKOLA 180
 Qy 181 LTLIDNRSRACHPCSPMKGSRGWESSEDCOSLRTVCAGGCAKCGPLPTDCCHEQC 240
 Db 181 LTLIDNRSRACHPCSPMKGSRGWESSEDCOSLRTVCAGGCAKCGPLPTDCCHEQC 240
 Qy 241 AAGTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 Db 241 AAGTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 Qy 301 YNYLSTDVGSCTVCPHNOEVTAEQDTCRCKSPCARVCYGLGMEHLREVRVAVTSAN 360
 Db 301 YNYLSTDVGSCTVCPHNOEVTAEQDTCRCKSPCARVCYGLGMEHLREVRVAVTSAN 360
 Qy 361 IQBFAGCKKIFGSLAFPEPSFDGDPASNTAPLOEQLOVFETLEEITGYLIVISAWPSLP 420
 Db 361 IQBFAGCKKIFGSLAFPEPSFDGDPASNTAPLOEQLOVFETLEEITGYLIVISAWPSLP 420
 Qy 421 DLSVFQNLQVIRGRIHLNGAYSILTQGLGISWGLRSLRSLGSLALIHNTLHLCFVHTV 480
 Db 421 DLSVFQNLQVIRGRIHLNGAYSILTQGLGISWGLRSLRSLGSLALIHNTLHLCFVHTV 480
 Qy 481 PWDQLFNPHQALLHTANRPEDECVGEGLAHQIARGHCHWGPGPTQVCNCSQPLRQBC 540
 Db 481 PWDQLFNPHQALLHTANRPEDECVGEGLAHQIARGHCHWGPGPTQVCNCSQPLRQBC 540
 Qy 541 VEECRVLQGLFREVYVNAHCLPCHPECPONGSVTCFPEADOCVCAHYKDDPPFCVARC 600
 Db 541 VEECRVLQGLFREVYVNAHCLPCHPECPONGSVTCFPEADOCVCAHYKDDPPFCVARC 600
 Qy 601 PSGVKPDLSTYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCFAEQRASPLTSIVSAVG 660
 Db 601 PSGVKPDLSTYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCFAEQRASPLTSIVSAVG 660
 Qy 661 ILLVVLGVVFGILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
 Db 661 ILLVVLGVVFGILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
 Qy 721 RKVKVLGSGAFGTWYGIWIPDGENVKI PVAIKVLRENTSPKANKETLDEAYVMAGVGP 780
 Db 721 RKVKVLGSGAFGTWYGIWIPDGENVKI PVAIKVLRENTSPKANKETLDEAYVMAGVGP 780
 Qy 781 YVSRLLIGICLTSTVQLVTQLMPTGYCLLDHVRENRLGSGQDLNWCQIAKMSYLEDDVR 840

Db 781 YVSRLLGICLTSTVQLVTQMPYGCCLLDHVRNRRGLSGQDLNWCMIKAGMSYLEDVR 840
Qy 841 LVHRDLAARNVLKSPNPKITDFGLARLLDIDETEHADGGKVPKIMWMALESILRRFT 900
Db 841 LVHRDLAARNVLKSPNPKITDFGLARLLDIDETEHADGGKVPKIMWMALESILRRFT 900
Qy 901 HQSDVWSYGVVWELMTGAKPYDGI PARETPDILLEKGERLPQPPICITIDVYIMVWKWM 960
Db 901 HQSDVWSYGVVWELMTGAKPYDGI PARETPDILLEKGERLPQPPICITIDVYIMVWKWM 960
Qy 961 IDSECRPRFRELVSFSESMARDPQRFVVIQNEDELGPASPLDSTFYRSILLEDDMGDLVDA 1020
Db 961 IDSECRPRFRELVSFSESMARDPQRFVVIQNEDELGPASPLDSTFYRSILLEDDMGDLVDA 1020
Qy 1021 EYLVPQGGFCPPDPAPGAGWVHRHRSSTRSGGDLTLGLPSEBEAPRPLARSEG 1080
Db 1021 EYLVPQGGFCPPDPAPGAGWVHRHRSSTRSGGDLTLGLPSEBEAPRPLARSEG 1080
Qy 1081 AGSDVFDGDLGWAAGLQSLPTDPSPLQRYSDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGWAAGLQSLPTDPSPLQRYSDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Qy 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTTPQ 1200
Db 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTTPQ 1200
Qy 1201 GGAAPOPHPPAFSPAFNLVYWDQPPERGAPPSFTFGTPTAENPEYGLDVPV 1255
Db 1201 GGAAPOPHPPAFSPAFNLVYWDQPPERGAPPSFTFGTPTAENPEYGLDVPV 1255

RESULT 5

AAW01111 standard; Protein; 1255 AA.

AC AAW01111;

DT 01-JAN-1997 (first entry)

DE HER-2/neu protein.

KW HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
KW breast cancer; ovary cancer; colon cancer; lung cancer;
KW prostate cancer; immunisation; tumour; vaccine; vector.

OS Homo sapiens.

FH Key Location/Qualifiers

FT Domain 676..1255

FT /label= Intracellular_domain

FT /note= "claimed domain, useful for immunisation"

XX WO9630514-A1.

XX 03-OCT-1996.

XX 28-MAR-1996; 96WO-US01689.

XX 31-MAR-1995; 95US-0414417.

XX (UNIW) UNIV WASHINGTON.

XX Cheever MA, Disis ML;

XX WPI; 1996-455361/45.

XX N-ESDB; AAT40739.

PT DNA encoding HER-2-neu poly-peptide(s) - used for prevention or
PT treatment of malignancies with which the HER-2/neu oncogene is
PT associated

XX Claim 2; Page 56-61; 71pp; English.

XX Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is
CC the product of the HER-2/neu oncogene (see also AAT40739). The
CC protein is over-expressed in various cancers, including breast,
CC ovarian, colon, lung and prostate. The intracellular domain of the
CC protein can be used to immunise an animal against a malignancy with
CC which the oncogene is associated. The polypeptide can be produced
CC in transformed host cells for use in immunisation. Alternatively,
CC animal calls are transfected in vivo or ex vivo with a viral vector
CC that directs expression of the polypeptide.

XX SQ Sequence 1255 AA;

Query Match 98.8%; Score 6729; DB 17; Length 1255;

Best Local Similarity 98.8%; Pred. No. 0;

Matches 1240; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLALLPPGAASQVCTGDMKRLPASPETHLDMLRHLVGGCQVQGNL 60

Db 1 MELAALCRWGLLALLPPGAASQVCTGDMKRLPASPETHLDMLRHLVGGCQVQGNL 60

Qy 61 ELYLPTNASLSFLQDIQEVQYVLI AHNQVRQVPLQLRLIVRGTLFEDNYALAVLNG 120

Db 61 ELYLPTNASLSFLQDIQEVQYVLI AHNQVRQVPLQLRLIVRGTLFEDNYALAVLNG 120

Qy 121 DPLNNTTPTVGASPGGLRELQLRSLTEILKGGVLIQRNPOLCYQDTILWKDI FHNQOLA 180

Db 121 DPLNNTTPTVGASPGGLRELQLRSLTEILKGGVLIQRNPOLCYQDTILWKDI FHNQOLA 180

Qy 181 LTLIDTNRSRACHPCSPMKGSRGWSSSDCQSLTRTVCAAGCARCKGLPTDCCHEQC 240

Db 181 LTLIDTNRSRACHPCSPMKGSRGWSSSDCQSLTRTVCAAGCARCKGLPTDCCHEQC 240

Qy 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Db 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy 301 YNYLSTDVGSCTLVCPLNQOEVTAEQCEKSKPCARVCYGLGMEHLREVRVTSAN 360

Db 301 YNYLSTDVGSCTLVCPLNQOEVTAEQCEKSKPCARVCYGLGMEHLREVRVTSAN 360

Qy 361 IQEFAGCKKIFGSLAFPLESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Db 361 IQEFAGCKKIFGSLAFPLESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLP 420

Qy 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLGSLALIHNNTHLCFVHTV 480

Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLGSLALIHNNTHLCFVHTV 480

Qy 481 PWDOLFRNPHQALLHTANRPEDECEVGRGLACHQLCARGHCWGPGPTQVCNCSQFLRGQEC 540

Db 481 PWDOLFRNPHQALLHTANRPEDECEVGRGLACHQLCARGHCWGPGPTQVCNCSQFLRGQEC 540

Qy 541 VEECRVLQGLPREYVYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFVCVARC 600

Db 541 VEECRVLQGLPREYVYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFVCVARC 600

Qy 601 PSGVKPDLSYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCAPAEORASPLTSISAVVG 660

Db 601 PSGVKPDLSYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCAPAEORASPLTSISAVVG 660

Qy 661 ILLVVLGVVFGILIKRRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720

Db 661 ILLVVLGVVFGILIKRRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720

Qy 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAGVGP 780

Db 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAGVGP 780

Qy 781 YVSRLLGICLTSTVQLVTQMPYGCCLLDHVRNRRGLSGQDLNWCMIKAGMSYLEDVR 840

Db 781 YVSRLLGICLTSTVQLVTQMPYGCCLLDHVRNRRGLSGQDLNWCMIKAGMSYLEDVR 840

Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWKMALESILRRFT 900
 Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWKMALESILRRFT 900
 Qy 901 HQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICITIDVYIMVKCWM 960
 Db 901 HQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICITIDVYIMVKCWM 960
 Qy 961 IDSECRPRFRELSEFSEFARMARDPQRFVVIQNEIDLGASPLDSTFYRSLLEDGMDLVA 1020
 Db 961 IDSECRPRFRELSEFSEFARMARDPQRFVVIQNEIDLGASPLDSTFYRSLLEDGMDLVA 1020
 Qy 1021 EYILVPOGFCPPAPGAGCAGWVHRHRSSTRSGGDLTLGLPSEBAPRSPASEG 1080
 Db 1021 EYILVPOGFCPPAPGAGCAGWVHRHRSSTRSGGDLTLGLPSEBAPRSPASEG 1080
 Qy 1081 AGSDVFDGLGMAAGLQSLPHTDPSPLORYSDPTVPLPSETDGYVAPLTCSPQPEYV 1140
 Db 1081 AGSDVFDGLGMAAGLQSLPHTDPSPLORYSDPTVPLPSETDGYVAPLTCSPQPEYV 1140
 Qy 1141 NQPDVREPQPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKOVFAFGGAVENPEYLTPQ 1200
 Db 1141 NQPDVREPQPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKOVFAFGGAVENPEYLTPQ 1200
 Qy 1201 GGAAPQHPHPPAFSPADNLYYWDQDPPERGAPSTFKGTPTAENPEYGLDVPV 1255
 Db 1201 GGAAPQHPHPPAFSPADNLYYWDQDPPERGAPSTFKGTPTAENPEYGLDVPV 1255

RESULT 6

AAW92406
 ID AAW92406 standard; Protein; 1255 AA.

AAW92406; .

XX 21-APR-1999 (first entry)

XX Human HER-2/neu oncogene protein.

XX HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;
 XX malignancy; treatment; tumour.

XX Homo sapiens.

XX Key Location/Qualifiers

XX Region 676..1255

XX /note= "region which elicits immune response"

XX US5869445-A.

XX 09-FEB-1999.

XX 01-APR-1996; 96US-0625101.

XX 01-APR-1996; 96US-0625101.

XX 17-MAR-1993; 93US-0033644.

XX 12-AUG-1993; 93US-0106112.

XX 31-MAR-1995; 95US-0414417.

XX (UNIW) UNIV WASHINGTON.

XX Cheever MA, Disis ML;

XX WPI; 1999-152835/13.

XX N-PSDB; AAX01912.

XX Use of HER-2/neu polypeptides - for eliciting an immune response to
 XX an HER-2/neu associated malignancy, particularly for treating or
 XX preventing tumours

XX Claim 3; Column 31-38; 26pp; English.

XX This sequence represents the human HER-2/neu oncogene protein. A fragment

CC of this protein is used in a method for eliciting or enhancing an immune
 CC response to HER-2/neu protein. The polypeptide can stimulate T cells and
 CC B cells to produce an immune response to the HER-2/neu protein. The
 CC method can be used for immunisation against a malignancy in which the
 CC HER-2/neu oncogene is associated and in the treatment of an existing
 CC tumour, or to prevent tumour occurrence or recurrence.

XX Sequence 1255 AA;

Query Match 98.8%; Score 6729; DB 20; Length 1255;
 Best Local Similarity 98.8%; Pred. No. 0;
 Matches 1240; Conservative. 3; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLALLPPGAASQVCTGTDMLRLPASPTHLDMLRHLYGQCQVQGNL 60
 Db 1 MELAALCRWGLLALLPPGAASQVCTGTDMLRLPASPTHLDMLRHLYGQCQVQGNL 60
 Qy 61 ELYLPTNASLSFLDIOEQVQYIKANSKFIGITELORLIRVGTQLFENYALAVLDNG 120
 Db 61 ELYLPTNASLSFLDIOEQVQYIKANSKFIGITELORLIRVGTQLFENYALAVLDNG 120
 Qy 121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNOLA 180
 Db 121 DPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNOLA 180
 Qy 181 LTLIDTNRSRACHPCSPMCKGSRGSESSDCQSLTRTVCCAGCARCKGPLPTDCCHEQC 240
 Db 181 LTLIDTNRSRACHPCSPMCKGSRGSESSDCQSLTRTVCCAGCARCKGPLPTDCCHEQC 240
 Qy 241 AAGCTGPKISDCLACLHFNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
 Db 241 AAGCTGPKISDCLACLHFNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
 Qy 301 YNYLSTDVGSCTLVCPLNQEVTAEDGTORCEKSPCARVCYGLGMEHLREVRVTSAN 360
 Db 301 YNYLSTDVGSCTLVCPLNQEVTAEDGTORCEKSPCARVCYGLGMEHLREVRVTSAN 360
 Qy 361 IQEFAGCKIIFGSLAFPLPESFDGDPASNTAPLOPEQLQVFETLEEITGYLIYISAWPDSL 420
 Db 361 IQEFAGCKIIFGSLAFPLPESFDGDPASNTAPLOPEQLQVFETLEEITGYLIYISAWPDSL 420
 Qy 421 DLSVFQNLQVIRIILHNGAYSILTLQGLIGISWILGRLSRLGSLALIHNTLHCFVHTV 480
 Db 421 DLSVFQNLQVIRIILHNGAYSILTLQGLIGISWILGRLSRLGSLALIHNTLHCFVHTV 480
 Qy 481 PWDLFRNPHQALLHTANRPEDECVGEGACHOLCARGHCWGPGTQCVCNCSQFLRGQEC 540
 Db 481 PWDLFRNPHQALLHTANRPEDECVGEGACHOLCARGHCWGPGTQCVCNCSQFLRGQEC 540
 Qy 541 VEECRVLQGLFREYVYVNAHCLPCHPECQONGSVTCFGEADQCVACAHYKDPFVCVARC 600
 Db 541 VEECRVLQGLFREYVYVNAHCLPCHPECQONGSVTCFGEADQCVACAHYKDPFVCVARC 600
 Qy 601 PSGVKPDLSPYIWKFPDSEGAQCPINCTHSCVDLDDKGCPAEQRASPLTISVAVG 660
 Db 601 PSGVKPDLSPYIWKFPDSEGAQCPINCTHSCVDLDDKGCPAEQRASPLTISVAVG 660
 Qy 661 ILLVVLGVVFGILIKRROOKIRKYTMRLLOTELVEPLTPSGAMPNQOMRILKETEL 720
 Db 661 ILLVVLGVVFGILIKRROOKIRKYTMRLLOTELVEPLTPSGAMPNQOMRILKETEL 720
 Qy 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKETLDEAYVMAGVGP 780
 Db 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKETLDEAYVMAGVGP 780
 Qy 781 YVSRLLGICLTSTVQLVLTQMPYGCLLDHVRENRLGSLQDLNWCWQIAKMSYLEDVR 840
 Db 781 YVSRLLGICLTSTVQLVLTQMPYGCLLDHVRENRLGSLQDLNWCWQIAKMSYLEDVR 840
 Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWKMALESILRRFT 900
 Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWKMALESILRRFT 900

Qy 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQPPICITIDVYMWKWM 960
 Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQPPICITIDVYMWKWM 960
 Qy 961 IDSECRPRFRELSEFSRMDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMGDLVDA 1020
 Db 961 IDSECRPRFRELSEFSRMDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMGDLVDA 1020
 Qy 1021 EYLVPQOGFFCPDPAFCAGGVHRRSSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1080
 Db 1021 EYLVPQOGFFCPDPAFCAGGVHRRSSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1080
 Qy 1081 AGSDVFDGLGWAAGKLOSLTHDPSPLOYSDDPTVPLPSETDGYVAPLTCSPQPEYV 1140
 Db 1081 AGSDVFDGLGWAAGKLOSLTHDPSPLOYSDDPTVPLPSETDGYVAPLTCSPQPEYV 1140
 Qy 1141 NQDVRPQPPSPREGPLPAARPAAGATLERAKTSLPGKNGVVKDVFAGGAVENPEYLTPQ 1200
 Db 1141 NQDVRPQPPSPREGPLPAARPAAGATLERAKTSLPGKNGVVKDVFAGGAVENPEYLTPQ 1200
 Qy 1201 GGAAPQPHPPAFSPAFDNLVYWDQPPPERGAPSTFKGTTAENPEYGLDVPV 1255
 Db 1201 GGAAPQPHPPAFSPAFDNLVYWDQPPPERGAPSTFKGTTAENPEYGLDVPV 1255

RESULT 7

AAB21198
 ID AAB21198 standard; protein; 1255 AA.

AC AAB21198;

XX 12-JAN-2001 (first entry)

DT Human HER-2/neu protein.

XX Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;
 KW breast cancer; prostate cancer; ovarian cancer; lung cancer;
 KW colon cancer.

OS Homo sapiens.

XX WO200044899-A1.

XX 03-AUG-2000.

XX 28-JAN-2000; 2000WO-US02164.

XX 29-JAN-1999; 99US-0117976.

XX (CORI-) CORIXA CORP.

PA (SMIK-) SMITHKLINE BEECHAM.

XX Cheever MA, Gheysen D;

PI WPI; 2000-505976/45.

XX N-P8DB; AAA89736.

PT HER-2/neu extracellular domain/phosphorylation domain fusion proteins
 PT useful for vaccinating against breast, ovarian, colon, lung and
 PT prostate cancers -

PS Claim 52; Fig 7; 128pp; English.

XX The present sequence is the human HER-2/neu protein. It is a member of
 CC the tyrosine kinase family of receptor-like glycoproteins and shows
 CC homology to the epidermal growth factor receptor (EGFR). It probably
 CC plays a part in cell growth and/or differentiation. The HER-2/neu
 CC gene is an oncogene. An HER-2/neu fusion protein comprising a
 CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
 CC domain may be used to treat or prevent cancer by eliciting or
 CC enhancing an immune response to the HER-2/neu protein. It may be used
 CC to treat malignancies such as breast, ovarian, colon, lung and
 CC prostate, cancers, and may be used as an antigen to vaccinate against

CC these neoplasias.
 XX
 SQ Sequence 1255 AA;

Query Match 98.8%; Score 6729; DB 21; Length 1255;
 Best Local Similarity 98.8%; Pred. No. 0;
 Matches 1240; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLALLPPGAASQVCTGTDMKLRLLPASPTHLDMLRHLHQGVVQGNL 60
 Db 1 MELAALCRWGLLALLPPGAASQVCTGTDMKLRLLPASPTHLDMLRHLHQGVVQGNL 60
 Qy 61 ELYVLPNASLSFLQDIQEVQVYIKANSKEFIGITELQRLIRVGTQLFEDNYALAVLDNG 120
 Db 61 ELYVLPNASLSFLQDIQEVQVYIKANSKEFIGITELQRLIRVGTQLFEDNYALAVLDNG 120
 Qy 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNQLA 180
 Db 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNQLA 180
 Qy 181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCOSLTRTVCCAGCARCKGPLPTDCCHEQC 240
 Db 181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCOSLTRTVCCAGCARCKGPLPTDCCHEQC 240
 Qy 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 Qy 301 YNYLSTDVGSCTLVCPLNHNEVTAEDGTQRCCKSPCARVCYGLGMEHLREVRVTSAN 360
 Db 301 YNYLSTDVGSCTLVCPLNHNEVTAEDGTQRCCKSPCARVCYGLGMEHLREVRVTSAN 360
 Qy 361 IQEFAGCKKIFGSLAFPEFDGDPASNTAPLOEQLOVFETLEEITGYLYISAWPSLP 420
 Db 361 IQEFAGCKKIFGSLAFPEFDGDPASNTAPLOEQLOVFETLEEITGYLYISAWPSLP 420
 Qy 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWLGRLSRLGSLALIHNTLHLCFVHTV 480
 Db 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWLGRLSRLGSLALIHNTLHLCFVHTV 480
 Qy 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQICARGHCWGPGTQCVNCSQPLRGQEC 540
 Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQICARGHCWGPGTQCVNCSQPLRGQEC 540
 Qy 541 VECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFCVARC 600
 Db 541 VECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFCVARC 600
 Qy 601 PSGVKPDLSYMPIWKFPDESGACQPCPINCTHSCVDLDDKGCAPAEQASPLTSISAVVG 660
 Db 601 PSGVKPDLSYMPIWKFPDESGACQPCPINCTHSCVDLDDKGCAPAEQASPLTSISAVVG 660
 Qy 661 ILLVVLGVVFGILIKRRQKIRKYTNRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
 Db 661 ILLVVLGVVFGILIKRRQKIRKYTNRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
 Qy 721 RKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMAGVGP 780
 Db 721 RKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMAGVGP 780
 Qy 781 YVSRLLGILCTSTVOLVTQMLPYGCLLDHVRNRLGRLGSDLLNWCMIKAGMSYLEDVR 840
 Db 781 YVSRLLGILCTSTVOLVTQMLPYGCLLDHVRNRLGRLGSDLLNWCMIKAGMSYLEDVR 840
 Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKWMALESILRRFT 900
 Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKWMALESILRRFT 900
 Qy 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQPPICITIDVYMWKWM 960
 Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQPPICITIDVYMWKWM 960
 Qy 961 IDSECRPRFRELSEFSRMDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMGDLVDA 1020

```

Db 961 IDSCRRFRRLVSEFSRMARDPQRFVVIQNEIDLGPAASPLDSTFYRSLEDDDDGDLVDA 1020
Qy 1021 BEYLVPOQGFCCPDPAFCAGGCMVHRHRSSTRSGGDLTLGLPSPSEEAAPRSLAPSEG 1080
Db 1021 BEYLVPOQGFCCPDPAFCAGGCMVHRHRSSTRSGGDLTLGLPSPSEEAAPRSLAPSEG 1080
Qy 1081 AGSDVFGDGLGMAAGKGLQSLPHTDPSPLQRYSDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFGDGLGMAAGKGLQSLPHTDPSPLQRYSDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Qy 1141 NQPDVRPQPPSPRGPLPAARPAGATLERAKTILSPGKNGVVKDYFAFGGAVENPEYLTPO 1200
Db 1141 NQPDVRPQPPSPRGPLPAARPAGATLERAKTILSPGKNGVVKDYFAFGGAVENPEYLTPO 1200
Qy 1201 GGAAPQHPHPPAFSPADNLYYWQDPPERGAPPTFKGTPTAENPEYLGLDVVP 1255
Db 1201 GGAAPQHPHPPAFSPADNLYYWQDPPERGAPPTFKGTPTAENPEYLGLDVVP 1255

RESULT 8
AAY84780
ID AAY84780 standard; Protein; 1255 AA.
XX
AC AAY84780;
XX
AC
XX
DT 08-AUG-2000 (first entry)
XX
DE Amino acid sequence of the SPLICE erbB-2 receptor protein.
XX
KW SPLICE erbB-2 receptor protein; cell transformation disorder; cancer;
KW tumor cell proliferation; tissue degeneration; arthropathy;
KW bone resorption; inflammatory disease; degenerative disorder;
KW wound healing.
XX
XX Homo sapiens.
XX
XX WO200020579-A1.
XX
XX 13-APR-2000.
XX
XX 01-OCT-1999; 99WO-CA00912.
XX
XX 02-OCT-1998; 98US-0165192.
XX
XX (UYMC-) UNIV MCMASTER.
XX
XX Muller WJ, Siegel PM;
XX
XX WPI; 2000-303768/26.
XX
XX N-PSDB; AAA14812.
XX
XX Nucleic acid encoding an erbB 2 receptor protein designated SPLICE
XX erbB-2, inhibitors of the protein are useful for treatment of cancer -
XX Claim 3; Fig 2; 60pp; English.
XX
XX The present sequence represents a SPLICE erbB-2 receptor protein. The
XX protein has an in-frame deletion of 16 amino acids, 2 of which are
XX conserved cysteine residues, compared to the unspliced protein. The
XX erbB-2 polynucleotide is used to construct probes for detecting
XX disorders of cell transformation such as cancer. Antibodies to the
XX protein may be used to detect SPLICE erbB-2 in a sample. Agents
XX (e.g. antisense oligonucleotides) which inhibit the expression of
XX SPLICE erbB-2 are useful for reducing tumor cell proliferation and
XX treating cancer. Substances which stimulate SPLICE erbB-2 are useful
XX for treating conditions involving damaged cells including conditions
XX in which degeneration of tissue occurs, such as arthropathy, bone
XX resorption, inflammatory diseases, degenerative disorders of the
XX central nervous system and wound healing.
XX
XX Sequence 1255 AA;

```

```

Query Match 98.8%; Score 6729; DB 21; Length 1255;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MEAAALCRWGLLALLPPGAASQVCTGTDMKRLPASPTHLDMLRHLVGGCCVVOGNL 60
Db 1 MEAAALCRWGLLALLPPGAASQVCTGTDMKRLPASPTHLDMLRHLVGGCCVVOGNL 60
Qy 61 ELTYLPTNASLSFLQDIQEVQYIKANSKFIKITELQRLRIVRGTLQFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIQEVQYIKANSKFIKITELQRLRIVRGTLQFEDNYALAVLDNG 120
Qy 121 DPLNTPPTVGTASPGGLREQLRLSLTEILKGVLIQRNPOLCYQDTTLWKDIFHKNOLA 180
Db 121 DPLNTPPTVGTASPGGLREQLRLSLTEILKGVLIQRNPOLCYQDTTLWKDIFHKNOLA 180
Qy 181 LTLIDTNRSRACHPCSPMKSGSCWGESSEDCSLTRTVCCAGGCARCKGLPTDCCHEQC 240
Db 181 LTLIDTNRSRACHPCSPMKSGSCWGESSEDCSLTRTVCCAGGCARCKGLPTDCCHEQC 240
Qy 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPLNHNOEVTABDGTQRCCKSPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNYLSTDVGSCTLVCPLNHNOEVTABDGTQRCCKSPCARVCYGLGMEHLREVRVTSAN 360
Qy 361 IQBFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFETLEETIGYLIYISAWPSLP 420
Db 361 IQBFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFETLEETIGYLIYISAWPSLP 420
Qy 421 DLSVFQNLQVIRGRIILHNGAYSILTLQGLGISWISGLRSLRELGLALIHNTLHCFVHTV 480
Db 421 DLSVFQNLQVIRGRIILHNGAYSILTLQGLGISWISGLRSLRELGLALIHNTLHCFVHTV 480
Qy 481 PWDQLFRNPHQALLHTANRPEDECEVGEGLACHOLCARGHCWGPGTQCVNCSQFLRQEC 540
Db 481 PWDQLFRNPHQALLHTANRPEDECEVGEGLACHOLCARGHCWGPGTQCVNCSQFLRQEC 540
Qy 541 VEESRVLOGLPREYVNAHCLPCHPECPONGSVTCFGEADOCVACAHYKDPFCVARC 600
Db 541 VEESRVLOGLPREYVNAHCLPCHPECPONGSVTCFGEADOCVACAHYKDPFCVARC 600
Qy 601 PSGVKPDLSPYMKPDEEGACQPCINCHSCVDLDDKGCPCAEQASPLTSISAVVG 660
Db 601 PSGVKPDLSPYMKPDEEGACQPCINCHSCVDLDDKGCPCAEQASPLTSISAVVG 660
Qy 661 ILLVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAOMRILKETEL 720
Db 661 ILLVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAOMRILKETEL 720
Qy 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVIRENTSPKANKETLDEAYVMAGVSP 780
Db 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVIRENTSPKANKETLDEAYVMAGVSP 780
Qy 781 YVSRLLGICLTSTVQLVTQLMYPGCLLDHVRNKRGLSGODLLNWCMIKAGMSYLEDVR 840
Db 781 YVSRLLGICLTSTVQLVTQLMYPGCLLDHVRNKRGLSGODLLNWCMIKAGMSYLEDVR 840
Qy 841 LVHRDLAARNVLKSPHNHVKITDFGLARLDDIDETEHADGGKVPIKWMALLESILRRFT 900
Db 841 LVHRDLAARNVLKSPHNHVKITDFGLARLDDIDETEHADGGKVPIKWMALLESILRRFT 900
Qy 901 HQSDVWSYGVYVWELMTFGAKPYDGIIPAREIPDLEKGERLPOPPCTIDVYIMWKCMW 960
Db 901 HQSDVWSYGVYVWELMTFGAKPYDGIIPAREIPDLEKGERLPOPPCTIDVYIMWKCMW 960
Qy 961 IDSECRPRFRELVSFSRMARDPQRFVVIQNEIDLGPAASPLDSTFYRSLEDDDDGDLVDA 1020
Db 961 IDSECRPRFRELVSFSRMARDPQRFVVIQNEIDLGPAASPLDSTFYRSLEDDDDGDLVDA 1020
Qy 1021 EYLVFPQOQFFCPDPAPFAGGCMVHRHRSSTRSGGDLTLGLPSPSEEAAPRSLAPSEG 1080

```



```
Db 1021 EEYLVPOQGFCCPDPAFCAGGVHRRSSSTRSGGDLTLGLEPSEEA PRSLAPSEG 1080
Qy 1081 AGSDVFDGDLGMGAAGLQSLPHTDPSPLQRYSDPTVPLPSETDGVVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMGAAGLQSLPHTDPSPLQRYSDPTVPLPSETDGVVAPLTCSPQPEYV 1140
Qy 1141 NOPDVRPQPSPRGPIPAARPAAGATLERAKTILSPGKNGVVKDYFAFGGAVENPEYLTPO 1200
Db 1141 NOPDVRPQPSPRGPIPAARPAAGATLERAKTILSPGKNGVVKDYFAFGGAVENPEYLTPO 1200
Qy 1201 GGAAPQHPHPAFPAFDNLYWDDPPRGAPSTFKGTPTAENPEYLGIDVFPV 1255
Db 1201 GGAAPQHPHPAFPAFDNLYWDDPPRGAPSTFKGTPTAENPEYLGIDVFPV 1255

RESULT 9
AAB85458
ID AAB85458 standard; Protein; 1255 AA.
XX
AC AAB85458;
XX
DT 25-SEP-2001 (first entry)
XX
DE Human HER-2/neu protein.
XX
KW Antigen-presenting cell; immunogenic; immune response; HER-2/neu;
KW oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.
XX
OS Homo sapiens.
XX
XX WO200153463-A2.
XX
PD 26-JUL-2001.
XX
XX 19-JAN-2001; 2001WO-US01850.
XX
XX 21-JAN-2000; 2000US-0177545.
XX
XX (CORI-) CORIXA CORP.
XX
XX Cheever MA, Hand-Zimmermann S;
XX
XX WPI; 2001-476112/51.
XX
XX N-PSDB; AAH23392.
XX
XX New antigen-presenting cells, useful as vaccines for eliciting or
XX enhancing an immune response to HER-2/neu protein, particularly useful
XX for treating or preventing cancer, e.g. breast cancer -
XX
XX Claim 2; Page 41-46; 49pp; English.
XX
XX The invention provides an isolated antigen-presenting cell, which
XX expresses at least an immunogenic portion of a polypeptide that produces
XX an immune response to HER-2/neu protein. The antigen-presenting cells are
XX useful as vaccines for eliciting or enhancing an immune response to
XX HER-2/neu protein, particularly in treating or preventing malignancies in
XX which the HER-2/neu oncogene is associated. Specifically, these are
XX useful for treating or preventing cancer, e.g. breast cancer, ovarian,
XX colon, lung or prostate cancers. The present sequence represents
XX the human HER-2/neu protein (also known as p185 or c-erbB2).
XX
XX Sequence 1255 AA;
XX
XX Query Match 98.8%; Score 6729; DB 22; Length 1255;
XX Best Local Similarity 98.8%; Pred. No. 0;
XX Matches 1240; Conservative 3; Mismatches 12; Indels 0; Gaps 0;
XX
XX 1 MELAAALCRWGLLALLPFGAASQTCTGDMKRLPASPETHLDMRLHYGCGVQGNL 60
XX 1 MELAAALCRWGLLALLPFGAASQTCTGDMKRLPASPETHLDMRLHYGCGVQGNL 60
XX
XX 61.ELTYLPTNASLSFLQDIEQVGVLIHNGVYVPLQRLRIVRGTQLFEDNYALAVLDNG 120
```

```
Db 61 ELTYLPTNASLSFLQDIEQVGVLIHNGVYVPLQRLRIVRGTQLFEDNYALAVLDNG 120
Qy 121 DPLNNTTPTVTCASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTTLWKDI PHKNOLA 180
Db 121 DPLNNTTPTVTCASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTTLWKDI PHKNOLA 180
Qy 181 LTLIDNRSRACHPCSPCKGSCWGESSEDCSLTRTVCCAGCARCKGLPTDCCHEQC 240
Db 181 LTLIDNRSRACHPCSPCKGSCWGESSEDCSLTRTVCCAGCARCKGLPTDCCHEQC 240
Qy 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPHNOEVTAEADGTQRCCKSKPCARVYGLGMEHREVRVAVTSAN 360
Db 301 YNYLSTDVGSCTLVCPHNOEVTAEADGTQRCCKSKPCARVYGLGMEHREVRVAVTSAN 360
Qy 361 IQEFAGCKIFGSLAFPLPESFDGDPASNTAPLOEQVFEETLEETGYLYISAWPDSL 420
Db 361 IQEFAGCKIFGSLAFPLPESFDGDPASNTAPLOEQVFEETLEETGYLYISAWPDSL 420
Qy 421 DLSVFQNLQVIRGRIHLHNGAYSLTQGLGISWGLSLRSLRELGSGLALIHNTLHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRIHLHNGAYSLTQGLGISWGLSLRSLRELGSGLALIHNTLHLCFVHTV 480
Qy 481 PWDQLFRNPHQALLHTANRDECEVGEGLACHQLCARGHCWGPGTQVCNCSQPLRQOEC 540
Db 481 PWDQLFRNPHQALLHTANRDECEVGEGLACHQLCARGHCWGPGTQVCNCSQPLRQOEC 540
Qy 541 VECRVLQGLPREYVNAHCLPCHPECOPOGNSVTCFGEADOCVACAHYKDPFVCARC 600
Db 541 VECRVLQGLPREYVNAHCLPCHPECOPOGNSVTCFGEADOCVACAHYKDPFVCARC 600
Qy 601 PSGVKPDLSTYMPWKPFDEEGACQPCPINCTHSCVDLDDKGC PAEQRASPLTSIISAVVG 660
Db 601 PSGVKPDLSTYMPWKPFDEEGACQPCPINCTHSCVDLDDKGC PAEQRASPLTSIISAVVG 660
Qy 661 ILLVVLGVVFGILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRLKETEL 720
Db 661 ILLVVLGVVFGILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRLKETEL 720
Qy 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMAGVGP 780
Db 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMAGVGP 780
Qy 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRLGSLQDLNWCMIAGMSYLEDDR 840
Db 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRLGSLQDLNWCMIAGMSYLEDDR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGGKVPDKWMALESILRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGGKVPDKWMALESILRRFT 900
Qy 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPICTIDVYMLVWKCM 960
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPICTIDVYMLVWKCM 960
Qy 961 IDSECRPRFRELSEFSEFMRMARDPQRFVVIQNEIDLGASPLDSTFYRSLLEDDDDGLVDA 1020
Db 961 IDSECRPRFRELSEFSEFMRMARDPQRFVVIQNEIDLGASPLDSTFYRSLLEDDDDGLVDA 1020
Qy 1021 EYILVPOQGFCCPDPAFCAGGVHRRSSSTRSGGDLTLGLEPSEEA PRSLAPSEG 1080
Db 1021 EYILVPOQGFCCPDPAFCAGGVHRRSSSTRSGGDLTLGLEPSEEA PRSLAPSEG 1080
Qy 1081 AGSDVFDGDLGMGAAGLQSLPHTDPSPLQRYSDPTVPLPSETDGVVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMGAAGLQSLPHTDPSPLQRYSDPTVPLPSETDGVVAPLTCSPQPEYV 1140
Qy 1141 NOPDVRPQPSPRGPIPAARPAAGATLERAKTILSPGKNGVVKDYFAFGGAVENPEYLTPO 1200
```

Db 1141 NQDVRPQPPSPREGPLPAARPAGATLERPKTSLPGKNGVVDYFAFGGAVENPEYLTPQ 1200
Qy 1201 GGAAAPQHPHPPAFPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYGLDVPV 1255
Db 1201 GGAAAPQHPHPPAFPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYGLDVPV 1255
RESULT 10
AAG88267
ID AAG88267 standard; Protein; 1255 AA.
XX
AC AAG88267;
XX
DT 11-SEP-2001 (first entry)
XX
DE HER2/neu amino acid sequence.
XX
KW Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
KW immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;
KW tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
XX
OS Homo sapiens.
XX
PN WO200141787-A1.
XX
PD 14-JUN-2001.
XX
PF 11-DEC-2000; 2000WO-US33591.
XX
PR 10-DEC-1999; 99US-0458299.
XX
PA (EPIM-) EPIMUNE INC.
XX
PI Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;
PI Keogh E;
XX
DR WPT; 2001-374995/39.
XX
PT An isolated prepared HER2/neu epitope useful in a vaccine for inducing
PT cellular immune responses for the prevention and treatment of cancer -
PS Disclosure; Page 15; 199pp; English.
XX
CC The present invention describes isolated prepared HER2/neu epitopes (I).
CC Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
CC culture in vitro and binds to a complex of an epitope (I), bound to a
CC human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)
CC and a second epitope and the peptide is less than 50 contiguous amino
CC acids that have 100% identity with a native peptide sequence of HER2/neu;
CC (3) a vaccine composition (III) comprising (II) and a pharmaceutical
CC excipient; (4) an isolated nucleic acid encoding a peptide comprising
CC (I); and (5) an isolated nucleic acid encoding (II). (I) has cytostatic
CC and immunostimulant activities, and can be used in vaccines. (I), (II)
CC and (III) are useful for inducing cellular immune responses for the
CC prevention and treatment of cancer. (I) and (II) are useful for
CC monitoring or evaluating an immune response to a tumour-associated
CC antigen when incubated with a T lymphocyte sample from a patient and
CC detecting the presence of bound T lymphocyte to (I) or (II). Epitope
CC based vaccines mean that immunosuppressive epitopes that may be present
CC in whole antigens may be avoided. Selected epitopes may be combined to
CC enhance immunogenicity. The possible pathological side effects caused by
CC infectious agents or whole protein antigen is eliminated. The vaccine
CC provides the ability to direct and focus an immune response to multiple
CC selected antigens from the same pathogen. Epitope-based anti-tumour
CC vaccines provides the opportunity to combine epitopes derived from
CC multiple tumour-associated molecules addressing the problem of tumour-
CC tumour variability and reducing the likelihood of tumour escape due to
CC antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
CC the exemplification of the present invention.
XX
SQ Sequence 1255 AA;

Query Match

98.8%; Score 6729; DB 22; Length 1255;

Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 3; Mismatches 12; Indels 0; Gaps 0;
Qy 1 MELAALCRWGLLALLPFGAASQVCTGTDMKRLPASPTHLDMLRHLRGCGVOVGNL 60
Db 1 MELAALCRWGLLALLPFGAASQVCTGTDMKRLPASPTHLDMLRHLRGCGVOVGNL 60
Qy 61 ELTYLPTNASLSFLQDIQEVQVYLIKANSKPIGITELORLRIVRGTOQLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIQEVQVYLIKANSKPIGITELORLRIVRGTOQLFEDNYALAVLDNG 120
Qy 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTTLWKDI FHNKOLA 180
Db 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTTLWKDI FHNKOLA 180
Qy 181 LTLIDTNRSRACHPCSPMKCKGSRGWESSDCOSLTRTVCAGGCARCKGLPTDCCHEQC 240
Db 181 LTLIDTNRSRACHPCSPMKCKGSRGWESSDCOSLTRTVCAGGCARCKGLPTDCCHEQC 240
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRTYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRTYTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPPLHNOEVTAEADGTQRCCKSPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNYLSTDVGSCTLVCPPLHNOEVTAEADGTQRCCKSPCARVCYGLGMEHLREVRVTSAN 360
Qy 361 IQEFAGCKKIFGSLAFPLPESFDGDPASNTAPLOEQLOVFETLEEITGYLYISAMPDLSL 420
Db 361 IQEFAGCKKIFGSLAFPLPESFDGDPASNTAPLOEQLOVFETLEEITGYLYISAMPDLSL 420
Qy 421 DLSVFQNLQVIRGRIILHNGAYSLTQGLGISWLGSLRSLRGLSLALIHNTLHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRIILHNGAYSLTQGLGISWLGSLRSLRGLSLALIHNTLHLCFVHTV 480
Qy 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQLCARGHCWGPGTQVCNCSOFLRQEC 540
Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQLCARGHCWGPGTQVCNCSOFLRQEC 540
Qy 541 VESCRVLQGLPREYVNAHCLPCHPFCOPONGSVTCFGEADOCVCAHVKDPPFCVARC 600
Db 541 VESCRVLQGLPREYVNAHCLPCHPFCOPONGSVTCFGEADOCVCAHVKDPPFCVARC 600
Qy 601 PSGVKPDLSTYMPKPFDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTISAVVG 660
Db 601 PSGVKPDLSTYMPKPFDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTISAVVG 660
Qy 661 ILLVVLGVVFGILIKRRQOKIRKYTMRLLOQTELVEPLTPSGAMPNQAMRILKETEL 720
Db 661 ILLVVLGVVFGILIKRRQOKIRKYTMRLLOQTELVEPLTPSGAMPNQAMRILKETEL 720
Qy 721 RKVKVLGSGAFGVYKGIWTPDGENVKIPVAIKVLRENTSPKANKETLDEAYVMAGVGP 780
Db 721 RKVKVLGSGAFGVYKGIWTPDGENVKIPVAIKVLRENTSPKANKETLDEAYVMAGVGP 780
Qy 781 YVSRLLGICLTSTVQLVTLQMPYGCLLDHVRENRLGSDODLLNWCNQIAKGSYLEDDVR 840
Db 781 YVSRLLGICLTSTVQLVTLQMPYGCLLDHVRENRLGSDODLLNWCNQIAKGSYLEDDVR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVP IKWMALESILRRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVP IKWMALESILRRRFT 900
Qy 901 HQSDVMSGYVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPPICTIDVYIMVKCWM 960
Db 901 HQSDVMSGYVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPPICTIDVYIMVKCWM 960
Qy 961 IDSECRPRFRELVSERFMRARDPQRFVWQNEIDGASPLDSTFYRSLLEDDMDGLVDA 1020
Db 961 IDSECRPRFRELVSERFMRARDPQRFVWQNEIDGASPLDSTFYRSLLEDDMDGLVDA 1020
Qy 1021 EYLVLPQQGFCDPDPAGAGVWHRRHSSTSGGGDLTLGLEPSEEPAPRSLAPSEG 1080
Db 1021 EYLVLPQQGFCDPDPAGAGVWHRRHSSTSGGGDLTLGLEPSEEPAPRSLAPSEG 1080

Db 1021 EYLVPQGFPCDPAPGAGGVHRRSSSTRSGGDLTLGLPSEEA PRSPLAPSEG 1080
 Qy 1081 AGSDVFDGDLGMAAKGLQSLPTDHPSPLOYSBDPTVPLPSETDGYVAPITCSPQPEYV 1140
 Db 1081 AGSDVFDGDLGMAAKGLQSLPTDHPSPLOYSBDPTVPLPSETDGYVAPITCSPQPEYV 1140
 Qy 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVVDVPAFGGAVENPEYLTPO 1200
 Db 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVVDVPAFGGAVENPEYLTPO 1200
 Qy 1201 GGAAPQHPHPPAFPAFNLIIYWDQDPPERGAPPSTFKGTPTAENPEYLGIDVPV 1255
 Db 1201 GGAAPQHPHPPAFPAFNLIIYWDQDPPERGAPPSTFKGTPTAENPEYLGIDVPV 1255

RESULT 11

AAE24067
 ID AAE24067 standard; Protein; 1255 AA.
 XX
 AC AAE24067;
 XX
 DT 23-SEP-2002 (first entry)
 XX
 DE Human Her-2 protein.
 XX
 KW Human; Her-2; epidermal growth factor receptor 2; infection; cancer;
 KW hyperproliferative disorder; prophylaxis; inflammation; antisense;
 KW tumour; gene therapy; phosphorothioate backbone.
 XX
 OS Homo sapiens.
 XX
 PN WO200222636-A1.
 XX
 PD 21-MAR-2002.
 XX
 PF 12-SEP-2001; 2001WO-US28572.
 XX
 PR 15-SEP-2000; 2000US-0663834.
 XX
 PA (ISIS-) ISIS PHARM INC.
 PI Bennett CF, Cowse LM;
 XX
 DR WPI; 2002-471192/50.
 DR N-FSDB; AAD38904.
 XX

Novel antisense oligonucleotide which modulates the expression of Human Epidermal Growth Factor receptor, Her2, is useful for treating tumors inflammation or to prevent infection in humans -

Example 13; Page 95-107; 116pp; English.

The invention relates to antisense compounds targeted to a nucleic acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2) that specifically hybridises with and inhibits the expression of Her2. Antisense compounds of the invention are used for treating diseases or conditions associated with Her2 such as hyperproliferative disorders e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary, neural or cardiac cancer. They are also useful prophylactically e.g. to prevent or delay infection, inflammation and tumour formation. The invention is also used in gene therapy. The present sequence is human Her-2 protein.

XX Sequence 1255 AA;

Query Match 98.8%; Score 6729; DB 23; Length 1255;

Best Local Similarity 98.8%; Pred. No. 0;

Matches 1240; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MELAACRWGILLALLPPGAASTQVCTGDMKRLPASPETHDMLRLHYGQCQVQGNL 60

Db 1 MELAACRWGILLALLPPGAASTQVCTGDMKRLPASPETHDMLRLHYGQCQVQGNL 60

Qy 61 ELTYLPTNASLSFLQDIOEQVQYIKANSKFIGITELQRLRIVRGTOLFEDNYALAVLDNG 120
 Db 61 ELTYLPTNASLSFLQDIOEQVQYIKANSKFIGITELQRLRIVRGTOLFEDNYALAVLDNG 120
 Qy 121 DPLNNTTPTVGTASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTTLWKDIFHKNQOLA 180
 Db 121 DPLNNTTPTVGTASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTTLWKDIFHKNQOLA 180
 Qy 181 LTLIDNRSRACHPCSPMKSGRCWGSSEDCOSLRTFVCAGGCARCKGLPTDCCHEQC 240
 Db 181 LTLIDNRSRACHPCSPMKSGRCWGSSEDCOSLRTFVCAGGCARCKGLPTDCCHEQC 240
 Qy 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTNTDTFESMPNPEGRYTFGASCVTACP 300
 Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTNTDTFESMPNPEGRYTFGASCVTACP 300
 Qy 301 YNYLSTDVGSCTTLVCPPLHNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRVTSAN 360
 Db 301 YNYLSTDVGSCTTLVCPPLHNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRVTSAN 360
 Qy 361 IOBFAGCKKIFGSLAFIPESFDGDPASNTAPLOEQLOVFETLEEITGYLYISAMPDPL 420
 Db 361 IOBFAGCKKIFGSLAFIPESFDGDPASNTAPLOEQLOVFETLEEITGYLYISAMPDPL 420
 Qy 421 DLSVFQNLQVIRGRILHNGAYSLTQGLGISWLSRLRELGLSLALIHNTLHLCFVHTV 480
 Db 421 DLSVFQNLQVIRGRILHNGAYSLTQGLGISWLSRLRELGLSLALIHNTLHLCFVHTV 480
 Qy 481 PWDQFRNPHQALLHTANRPEDECVGEGLAHQICARGHCGWPGPTQCVNCSQFLRGQEC 540
 Db 481 PWDQFRNPHQALLHTANRPEDECVGEGLAHQICARGHCGWPGPTQCVNCSQFLRGQEC 540
 Qy 541 VEECRVLQGLPREYVNAHCLPCHPECOQONGSVTCFGEADQCVACAHYKDPFCVARC 600
 Db 541 VEECRVLQGLPREYVNAHCLPCHPECOQONGSVTCFGEADQCVACAHYKDPFCVARC 600
 Qy 601 PSGVKPDLSTYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660
 Db 601 PSGVKPDLSTYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCPAEQRASPLTSIVSAVVG 660
 Qy 661 ILLVVLGVVFGILIKRRQKIRKYTWRLLEQTELVEPLTPSGAMPNQAMRILKETEL 720
 Db 661 ILLVVLGVVFGILIKRRQKIRKYTWRLLEQTELVEPLTPSGAMPNQAMRILKETEL 720
 Qy 721 RKVKVLGSGAGTGYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVWAGVSP 780
 Db 721 RKVKVLGSGAGTGYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVWAGVSP 780
 Qy 781 YVSRLLGICLTSTVOLVTQMLPYGCLLDHVRENRGLGSODLLNWCQIAKMSYLEDVR 840
 Db 781 YVSRLLGICLTSTVOLVTQMLPYGCLLDHVRENRGLGSODLLNWCQIAKMSYLEDVR 840
 Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLIDETEHADGGKVPKIKMALESILRRFT 900
 Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLIDETEHADGGKVPKIKMALESILRRFT 900
 Qy 901 HQSDVWSYGVTVWELMTFGAPYDGIIPAREIPDLLEKGERLPQPPICITDVIYMIWKCWM 960
 Db 901 HQSDVWSYGVTVWELMTFGAPYDGIIPAREIPDLLEKGERLPQPPICITDVIYMIWKCWM 960
 Qy 961 IDSECRPRFRELVSFERSMARDPQFVVIQNEIDLGPASPLDSTFYRSLLEDDDDMGDLVDA 1020
 Db 961 IDSECRPRFRELVSFERSMARDPQFVVIQNEIDLGPASPLDSTFYRSLLEDDDDMGDLVDA 1020
 Qy 1021 EYLVLPQGGFFCPDPAFCAGGMVHRRSSSTRSGGDLTLGLPSEEA PRSPLAPSEG 1080
 Db 1021 EYLVLPQGGFFCPDPAFCAGGMVHRRSSSTRSGGDLTLGLPSEEA PRSPLAPSEG 1080
 Qy 1081 AGSDVFDGDLGMAAKGLQSLPTDHPSPLOYSBDPTVPLPSETDGYVAPITCSPQPEYV 1140
 Db 1081 AGSDVFDGDLGMAAKGLQSLPTDHPSPLOYSBDPTVPLPSETDGYVAPITCSPQPEYV 1140
 Qy 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVVDVPAFGGAVENPEYLTPO 1200

Db 1141 NQDVRPQPPRSGPLPAAPAGATLPRKTLSPGKNGVVKVDFAFGAVENPEYLTPO 1200
Qy 1201 GGAAPQHPHPPAFPAFNLWYDQDPPRGAPSTFKGTPTAENPEYLGIDVPV 1255
Db 1201 GGAAPQHPHPPAFPAFNLWYDQDPPRGAPSTFKGTPTAENPEYLGIDVPV 1255
RESULT 12
AAE20479
ID AAE20479 standard; Protein; 1255 AA.
XX AAE20479;
AC AAE20479;
DT 01-JUL-2002 (first entry)
DE Human Her-2/neu protein.
KW Human; Her-2/Neu protein; immune response; gene therapy; breast cancer;
KW human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.
XX OS Homo sapiens.
XX FH Location/Qualifiers
FT Region 1021..1030
FT /note= "Naturally processed HLA-B44-restricted epitope"
XX PN WO200214503-A2.
XX PD 21-FEB-2002.
XX PF 14-AUG-2001; 2001WO-US41733.
XX PR 14-AUG-2000; 2000US-225152P.
XX PR 28-SEP-2000; 2000US-236428P.
XX PR 21-FEB-2001; 2001US-270520P.
XX PA (CORI-) CORIXA CORP.
XX PI Hand-zimmermann S, Cheever MA, Foy TM, Lodes MJ, Kalos MD;
PI McNeill PD, Vedwick TS;
XX WPI; 2002-280758/32.
XX DR N-PSDB; AAD32743.
XX PT Novel isolated Her-2/Neu polypeptide composition useful for therapy,
XX prevention and diagnosis of cancer, preferably breast cancer
XX Disclosure; Page 114-117; 129pp; English.
XX The invention relates to an isolated Her-2/Neu polypeptide composition
XX effective for eliciting an immune response. The invention is useful for
XX eliciting an immune response in a patient, where the patient is human
XX leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.
XX The composition is useful for the therapy and diagnosis of cancer,
XX preferably breast cancer, in pharmaceutical compositions, e.g., vaccine
XX and other compositions for the diagnosis, prevention and treatment of
XX human malignancies, for stimulating and/or expanding T cells specific for
XX Her-2/Neu polypeptide and for inhibiting the development of cancer in a
XX patient. The invention is useful for stimulating a T cell response in a
XX human patient, as probe or primer for nucleic acid hybridisation, to
XX selectively form duplex molecules with complementary stretches of the
XX entire Her-2/Neu gene or gene fragments of interest, to isolate a full
XX length gene from a suitable library, and to direct expression of a
XX polypeptide in appropriate host cells. The composition is useful in
XX prophylactic or therapeutic applications and for the treatment of cancer,
XX preferably for the immunotherapy of breast cancer and other Her-2/Neu-
XX associated malignancies. The invention is useful in gene therapy. The
XX present sequence is human Her-2/neu protein.
XX Sequence 1255 AA;

Query Match

98.8%; Score 6729; DB 23; Length 1255;

Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 3; Mismatches 12; Indels 0; Gaps 0;
Qy 1 MELAALCRWGLLALLPPGAASQVCTGTDKMLRASPETHLDMLRHLVGCQVQGNL 60
Db 1 MELAALCRWGLLALLPPGAASQVCTGTDKMLRASPETHLDMLRHLVGCQVQGNL 60
Qy 61 ELYLPTNASLSFLQDIOEVQVYKANSKFITELQRLRIVRGTOQLFEDNYALAVLDNG 120
Db 61 ELYLPTNASLSFLQDIOEVQVYKANSKFITELQRLRIVRGTOQLFEDNYALAVLDNG 120
Qy 121 DPLNNTTPTVTCASPGGLRELQSLTEILKGGVLIQRNPOLCYQDTILWKDIFHNQOLA 180
Db 121 DPLNNTTPTVTCASPGGLRELQSLTEILKGGVLIQRNPOLCYQDTILWKDIFHNQOLA 180
Qy 181 LTLIDTNRSRACHPCSPMKCKGRCWGSSEDCQSLTRTVCAAGGCARCKGPLPTDCCHEQC 240
Db 181 LTLIDTNRSRACHPCSPMKCKGRCWGSSEDCQSLTRTVCAAGGCARCKGPLPTDCCHEQC 240
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPLNHNEVTAEDGTQCEKCKSKPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNYLSTDVGSCTLVCPLNHNEVTAEDGTQCEKCKSKPCARVCYGLGMEHLREVRVTSAN 360
Qy 361 IQEPAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFETLEETGYLYISAWPDSLP 420
Db 361 IQEPAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFETLEETGYLYISAWPDSLP 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSLTLOGLGISWGLRSLRELGLSLALIHNTHLFCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSLTLOGLGISWGLRSLRELGLSLALIHNTHLFCFVHTV 480
Qy 481 PWDQLFRNPHQALHTANRPEDECVGEGLAHQICARGHCHGPGPTQCVNCSQFLRGQEC 540
Db 481 PWDQLFRNPHQALHTANRPEDECVGEGLAHQICARGHCHGPGPTQCVNCSQFLRGQEC 540
Qy 541 VECEVLQGLPREYVNAHCLPCHPEQOPNGSVTCFPGPEADQCVACAHYKDPFCVARC 600
Db 541 VECEVLQGLPREYVNAHCLPCHPEQOPNGSVTCFPGPEADQCVACAHYKDPFCVARC 600
Qy 601 PSGVKPDLSYNPIWKFPDEBGACQPCPINCTHSCVDLDDKGCAPAEQASPLTISVAVG 660
Db 601 PSGVKPDLSYNPIWKFPDEBGACQPCPINCTHSCVDLDDKGCAPAEQASPLTISVAVG 660
Qy 661 ILLVVLGVVFGILIKRRQOKIRKYTNRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
Db 661 ILLVVLGVVFGILIKRRQOKIRKYTNRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
Qy 721 RKVKVLGSGAFVYKGIWIPDGENVKIPVAIKVIRENTSPKANKETLDEAYVWAGVSP 780
Db 721 RKVKVLGSGAFVYKGIWIPDGENVKIPVAIKVIRENTSPKANKETLDEAYVWAGVSP 780
Qy 781 YVSRLLGICLTSTVQLVTLQMPYGCILLDHVRENKRLGSDLLNWCQIAKMSYLEBVR 840
Db 781 YVSRLLGICLTSTVQLVTLQMPYGCILLDHVRENKRLGSDLLNWCQIAKMSYLEBVR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKWMALESILRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKWMALESILRRFT 900
Qy 901 HQSDVMSYGVTVWELMTFGAPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVKCW 960
Db 901 HQSDVMSYGVTVWELMTFGAPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVKCW 960
Qy 961 IDSECRPRFRELVSFERSMARDPQRFVIONEDLGPASPLDSTFYRSLLEDGMDGLVDA 1020
Db 961 IDSECRPRFRELVSFERSMARDPQRFVIONEDLGPASPLDSTFYRSLLEDGMDGLVDA 1020
Qy 1021 EYLVLPQQGFCDPAPGAGMVHHRHSSTRSGGDLTLGLEPSEEPAPRSLPASEG 1080

Db 1021 BEYLVPOQFFCPDPAPGAGGVHRRSSSTRSGGDLTLGLPSEBEEAPRSLAPSEG 1080
 Qy 1081 AGSDVFDGDLGMAKGLQSLPTDHPSPLOQRYSEDPTVPLPSETDGYVAPITCSPQBEYV 1140
 Db 1081 AGSDVFDGDLGMAKGLQSLPTDHPSPLOQRYSEDPTVPLPSETDGYVAPITCSPQBEYV 1140
 Qy 1141 NOPDVRPQPPSPRGPLPAARPAAGATLERAKTLPQKNGVVKDVFAGGAVENPEYILTPQ 1200
 Db 1141 NOPDVRPQPPSPRGPLPAARPAAGATLERAKTLPQKNGVVKDVFAGGAVENPEYILTPQ 1200
 Qy 1201 GGAAPQHPHPPAFSPADNLNLYWQDPPERGAPPSTFKGTAEENPEYLGDDVPV 1255
 Db 1201 GGAAPQHPHPPAFSPADNLNLYWQDPPERGAPPSTFKGTAEENPEYLGDDVPV 1255

RESULT 13

AAM51143
 ID AAM51143 standard; Protein; 1255 AA.

AC AAM51143;

XX 17-JUN-2002 (first entry)

XX Human Her-2/neu oncogene-encoded p185 glycoprotein.

XX Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;
 KW tyrosine kinase; receptor; c-erbB2; gene therapy.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Domain 1..653 "extracellular domain"

FT Domain 676..1255

FT Domain "intracellular domain"

FT Domain 990..1255

FT Domain "phosphorylation domain"

XX WO200212341-A2.

XX 14-FEB-2002.

XX 03-AUG-2001; 2001WO-US24283.

XX 03-AUG-2000; 2000US-0632507.

XX (CORI-) CORIXA CORP.

XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX Cheever MA, Gheysen D;

XX WPI; 2002-241743/29.

XX N-PSDB; ABA92250.

XX Her-2/neu fusion protein for treating or preventing cancer by eliciting
 PT or enhancing an immune response to the protein, has Her-2/neu
 PT extracellular domain fused to Her-2/neu intracellular or
 PT phosphorylation domain

XX Claim 68; Fig 7; 141pp; English.

XX The present sequence is that of human Her-2/neu (p185 glycoprotein
 CC or c-erbB2), an oncogenic self-protein and target for anti-cancer
 CC vaccines. The Her-2/neu gene is amplified and p185 is overexpressed
 CC in a variety of cancers, including breast, ovarian, colon, lung and
 CC prostate cancer. Her-2/neu is a member of the tyrosine kinase
 CC family of receptor-like glycoproteins. It comprises an extracellular
 CC domain with homology to the epidermal growth factor receptor
 CC (EGFR), a highly hydrophobic transmembrane domain and a C-terminal
 CC intracellular domain that also shows homology to EGFR. Its
 CC overexpression correlates with a poor prognosis in breast and
 CC ovarian cancers. The invention provides Her-2/neu fusion
 CC proteins, nucleic acids encoding them, viral vectors, and vaccines

CC comprising the fusion proteins or nucleic acid molecules. In
 CC preferred fusion proteins, the extracellular domain of a Her-2/neu
 CC protein is fused to a Her-2/neu intracellular domain or
 CC phosphorylation domain (or its bclapd fragment). An immune
 CC response to Her-2/neu protein is elicited or enhanced by
 CC administering the fusion protein in the form of a vaccine, or by
 CC transfecting cells of an animal ex vivo with a nucleic acid
 CC encoding the fusion protein, and delivering the transfected cells
 CC to the animal. The fusion proteins, nucleic acids, and isolated
 CC specific T-cells are useful for inhibiting the development of a
 CC cancer, especially breast, ovarian, colon, lung or prostate cancer
 CC in a patient. T cells that specifically react with a Her-2/neu
 CC fusion protein can be used to remove tumour cells from a sample in
 CC order to inhibit the development of cancer in a patient.

XX Sequence 1255 AA;

Query Match 98.8%; Score 6729; DB 23; Length 1255;
 Best Local Similarity 98.8%; Pred. No. 0;
 Matches 1240; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLALLPFGAASSTQVCTGDMKRLPASPETHLDMRLHLYGCGVQGNL 60
 Db 1 MELAALCRWGLLALLPFGAASSTQVCTGDMKRLPASPETHLDMRLHLYGCGVQGNL 60
 Qy 61 ELTYLPTNASISFLQDIQEVQYIKANSKFTIGITELQRLRIVRGTLQFEDNYALVLDNG 120
 Db 61 ELTYLPTNASISFLQDIQEVQYIKANSKFTIGITELQRLRIVRGTLQFEDNYALVLDNG 120
 Qy 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHNQOLA 180
 Db 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHNQOLA 180
 Qy 181 LTLIDNRSRACHPCSPMCKGSRGWGSSSDCQSLRTVTCAGGCARCKGPIPTCCCHQC 240
 Db 181 LTLIDNRSRACHPCSPMCKGSRGWGSSSDCQSLRTVTCAGGCARCKGPIPTCCCHQC 240
 Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 Qy 301 YNYLSTDVGSCITLVCPLNHNOEVTAEADGTQRCCKSKPCARVCYGLGMEHLREVAVTSAN 360
 Db 301 YNYLSTDVGSCITLVCPLNHNOEVTAEADGTQRCCKSKPCARVCYGLGMEHLREVAVTSAN 360
 Qy 361 IQEPAGCKKIFGSLAFLPESFDGDPASNTAPLQEQLOVFTLEETITGLYISAWPDSLP 420
 Db 361 IQEPAGCKKIFGSLAFLPESFDGDPASNTAPLQEQLOVFTLEETITGLYISAWPDSLP 420
 Qy 421 DLSVFQNLQVIRGRILHNGAYSLSLTQGLGISWLGRLSRLRELGLALIHNNHLCFVHTV 480
 Db 421 DLSVFQNLQVIRGRILHNGAYSLSLTQGLGISWLGRLSRLRELGLALIHNNHLCFVHTV 480
 Qy 481 PWDOLFNPQHALLHTANRPEDECEVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
 Db 481 PWDOLFNPQHALLHTANRPEDECEVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
 Qy 541 VEECRVLQGLPREYVYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVARC 600
 Db 541 VEECRVLQGLPREYVYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVARC 600
 Qy 601 PSGVKPDLNLYWPIWKFPDEGACQPCINCTHSCVDLDDKGCAPAEORASPLTISVAVVG 660
 Db 601 PSGVKPDLNLYWPIWKFPDEGACQPCINCTHSCVDLDDKGCAPAEORASPLTISVAVVG 660
 Qy 661 ILLVVLGVVFGIILIKRQKIRKYTMRRLQETELVEPLTPSGAMPNQAQMRILKETEL 720
 Db 661 ILLVVLGVVFGIILIKRQKIRKYTMRRLQETELVEPLTPSGAMPNQAQMRILKETEL 720
 Qy 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVAVGVGSP 780
 Db 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVAVGVGSP 780

QY 781 YVSRLLGICLTSTVQLVLTQMPYGLDHRVNRGRGLSGDILLNWCMIKAGMSYLEVDV 840
 Db 781 YVSRLLGICLTSTVQLVLTQMPYGLDHRVNRGRGLSGDILLNWCMIKAGMSYLEVDV 840
 QY 841 LVHRDLAARNVLKSPNVKLTDFGLARLLDIDETEHADGKVPKWKMALESILRRFT 900
 Db 841 LVHRDLAARNVLKSPNVKLTDFGLARLLDIDETEHADGKVPKWKMALESILRRFT 900
 QY 901 HQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPICTIDVYIMVKCWM 960
 Db 901 HQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPICTIDVYIMVKCWM 960
 QY 961 IDSECRPRFRELVSFSEMRARDPQRFVVIQNEIDGPASPLDSTFYRSLLEDDMGDLVDA 1020
 Db 961 IDSECRPRFRELVSFSEMRARDPQRFVVIQNEIDGPASPLDSTFYRSLLEDDMGDLVDA 1020
 QY 1021 EYLVPQGGFCFDPAPGAGGVHHRSSSTRSGGDLTLGLEPSEBEAPRSLAPSEG 1080
 Db 1021 EYLVPQGGFCFDPAPGAGGVHHRSSSTRSGGDLTLGLEPSEBEAPRSLAPSEG 1080
 QY 1081 AGSDVFDGDLGWAAGKLSLTPHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
 Db 1081 AGSDVFDGDLGWAAGKLSLTPHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
 QY 1141 NQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVDVFAFGAVENPEYLTPO 1200
 Db 1141 NQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVDVFAFGAVENPEYLTPO 1200
 QY 1201 GGAAPQHPHPPAFSPAFNLVYWDODPPERGAPPSTFKGTPTAENPEYLGDDVPV 1255
 Db 1201 GGAAPQHPHPPAFSPAFNLVYWDODPPERGAPPSTFKGTPTAENPEYLGDDVPV 1255
 RESULT 14
 ID AAU77114 standard; Protein; 1255 AA.
 XX AAU77114;
 AC AAU77114;
 DT 05-JUN-2002 (first entry)
 XX Human Her-2/neu polypeptide.
 DE Human; Her-2/neu; cytostatic; haematological malignancy; CML;
 KW acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;
 KW chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;
 KW Hodgkin's lymphoma; T cell therapy.
 XX Homo sapiens.
 OS WO200213847-A2.
 FN 21-FEB-2002.
 PD 13-AUG-2001; 2001WO-US25408.
 XX 14-AUG-2000; 2000US-0638280.
 PR 28-SEP-2000; 2000US-0675904.
 XX (CORI-) CORIXA CORP.
 PA Gaiger A, Cheever MA, Hand-zimmermann S;
 XX WPI; 2002-280741/32.
 DR N-PSDB; ABK10730.
 DR Inhibiting haematological malignancy development by administering
 XX polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
 PT encoding the polypeptide, or antigen presenting cells expressing the
 PT polypeptide.
 XX Disclosure; Page 71-74; 74pp; English.
 PS
 XX

CC The invention relates to a method for inhibiting development of
 CC haematological malignancy in a patient by administering a polypeptide
 CC comprising an immunogenic portion of Her-2/neu or a polynucleotide
 CC encoding the polypeptide. Antigen presenting cells that express the
 CC protein can also be administered. The sequences are used for inhibiting
 CC development of haematological malignancy such as acute myelogenous
 CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
 CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
 CC lymphoma. This sequence represents the human Her-2/neu polypeptide.
 XX

SQ Sequence 1255 AA;

Query Match 98.8%; Score 6729; DB 23; Length 1255;
 Best Local Similarity 98.8%; Pred. No. 0;
 Matches 1240; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLALLPPGAATQVCTGDKMLRASPETHLDMLRHLVGGCQVQGNL 60
 Db 1 MELAALCRWGLLALLPPGAATQVCTGDKMLRASPETHLDMLRHLVGGCQVQGNL 60
 QY 61 ELTYLPTNASLSFLQDIQEVQYIKANSKFTIGITELQRLRIVRGTLQFEDNYALAVLDNG 120
 Db 61 ELTYLPTNASLSFLQDIQEVQYVLIHNNQVRVPLQLRLIVRGTLQFEDNYALAVLDNG 120
 QY 121 DPLNNTTPTVGTASPGGLREILQLRSLTEILKGGVLIQRNPOLCYQDTILWKDIFPHKNOLA 180
 Db 121 DPLNNTTPTVGTASPGGLREILQLRSLTEILKGGVLIQRNPOLCYQDTILWKDIFPHKNOLA 180
 QY 181 LTLIDTNRACHPCSPMKSGRCWGESSEDCQSLRTVTCAGGCARCKGPLPTCCCHQC 240
 Db 181 LTLIDTNRACHPCSPMKSGRCWGESSEDCQSLRTVTCAGGCARCKGPLPTCCCHQC 240
 QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 QY 301 YNYLSTDVGSCTLVCPLNHNEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVAVTSAN 360
 Db 301 YNYLSTDVGSCTLVCPLNHNEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVAVTSAN 360
 QY 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFETLEETGYLIYSAMPDPLP 420
 Db 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFETLEETGYLIYSAMPDPLP 420
 QY 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLGSLALIHNNTHLCFVHTV 480
 Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLGSLALIHNNTHLCFVHTV 480
 QY 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
 Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
 QY 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFPGPADQCVCAHYKDPFCVARC 600
 Db 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFPGPADQCVCAHYKDPFCVARC 600
 QY 601 PSGVKPDLSYNPIWKFPDEEGACQPCPINCTHSCVDLDDKGCFAEQRASPLTSTVSAVVG 660
 Db 601 PSGVKPDLSYNPIWKFPDEEGACQPCPINCTHSCVDLDDKGCFAEQRASPLTSTVSAVVG 660
 QY 661 ILLVVVLGVVFGILIKRRQKIRKYTWRRLLQETELVEPLTPSGAMPNQMRILKETEL 720
 Db 661 ILLVVVLGVVFGILIKRRQKIRKYTWRRLLQETELVEPLTPSGAMPNQMRILKETEL 720
 QY 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETILDEAYVMAGVGP 780
 Db 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETILDEAYVMAGVGP 780
 QY 781 YVSRLLGICLTSTVQLVLTQMPYGLDHRVNRGRGLSGDILLNWCMIKAGMSYLEVDV 840
 Db 781 YVSRLLGICLTSTVQLVLTQMPYGLDHRVNRGRGLSGDILLNWCMIKAGMSYLEVDV 840
 QY 841 LVHRDLAARNVLKSPNVKLTDFGLARLLDIDETEHADGKVPKWKMALESILRRFT 900

841 LVHRDLAARNVVKSPNHNKIDTGLARLLDIDETEHADGKVPKWMALLESILRRFT 900
 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYMIWVKCWM 960
 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYMIWVKCWM 960
 961 IDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSILLEDDMGDLVDA 1020
 961 IDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSILLEDDMGDLVDA 1020
 1021 BEYLVPQGGFFCPCPAPGAGGMVHRRSSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1080
 1021 BEYLVPQGGFFCPCPAPGAGGMVHRRSSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1080
 1081 AGSDVFDGDLGNGAAKGLQSLPTHDPSPLQRYSDPTVPLPSETDGYVAPLTCSPQPEYV 1140
 1081 AGSDVFDGDLGNGAAKGLQSLPTHDPSPLQRYSDPTVPLPSETDGYVAPLTCSPQPEYV 1140
 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPO 1200
 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLTPO 1200
 1201 GGAAPQHPHPPAFSPAFNLYYWDQDPPERGAPSTFKGTPTAENPEYIGLDVPV 1255
 1201 GGAAPQHPHPPAFSPAFNLYYWDQDPPERGAPSTFKGTPTAENPEYIGLDVPV 1255

RESULT 15

AAR39568

ID AAR39568 standard; Protein; 1433 AA.

XX AAR39568;

XX

XX

XX 07-FEB-1994 (first entry)

XX

XX Sequence of c-erbB-2 tumour antigen.

XX

XX Tumour antigen; c-erbB-2; glycoprotein.

XX

XX Homo sapiens.

XX

XX WO9316185-A.

XX

XX 19-AUG-1993.

XX

XX 05-FEB-1993; 93WO-US01055.

XX

XX 06-FEB-1992; 92US-0831967.

XX

XX (CETU) CETUS ONCOLOGY CORP.

XX (CREA-) CREATIVE BIOMOLECULES INC.

XX

XX Houston LL, Huston JS, Oppermann H, Ring DB;

XX

XX WPI; 1993-272889/34.

XX

XX N-PSDB; AAQ46083.

XX

XX New single chain Fv polypeptide binding to C-erbB-2 tumour

XX antigen - for imaging or treating breast or ovarian cancer etc.

XX

XX Disclosure; pages 48-54; 87pp; English.

XX

XX c-erbB-2 refers to a protein antigen expressed on the surface of

XX tumour cells, such as breast and ovarian tumour cells, which is an

XX approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric

XX pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39583 represents

XX the location of a stop codon in AAQ46083.

XX

XX Sequence 1433 AA;

XX

XX Query Match 98.2%; Score 6686; DB 14; Length 1433;

XX Best Local Similarity 98.2%; Pred. No. 0;

XX

	Matches 1233;	Conservative	5;	Mismatches	17;	Indels	0;	Gaps	0;
Qy	1	MELAAALCRWGLLALPPGAAS	TOVCTG	TDMKLR	LPAS	PETHLDM	LRLH	YGOC	VOVGNL 60
Db	1	MELAAALCRWGLLALPPGAAS	TOVCTG	TDMKLR	LPAS	PETHLDM	LRLH	YGOC	VOVGNL 60
Qy	61	ELTYLPTNASLSFLQDIQEVQ	QYIKANS	KFQITEL	QRLRI	VRGTOL	FEDNYA	LAVLDNG 120	
Db	61	ELTYLPTNASLSFLQDIQEVQ	QYIAHNV	QVQVLP	QRLRI	VRGTOL	FEDNYA	LAVLDNG 120	
Qy	121	DPLNTPPTVGA	SPGGLREL	QLSLTB	ILKGGV	LIQRN	POLCYOD	TTLWKD	IFHKNOLA 180
Db	121	DPLNTPPTVGA	SPGGLREL	QLSLTB	ILKGGV	LIQRN	POLCYOD	TTLWKD	IFHKNOLA 180
Qy	181	LTLDITNRSACHPC	SMCKGSC	WGSESS	DCOSL	TRTV	CAGG	CARCKG	PLPTDCCHEQC 240
Db	181	LTLDITNRSACHPC	SMCKGSC	WGSESS	DCOSL	TRTV	CAGG	CARCKG	PLPTDCCHEQC 240
Qy	241	AAGCTGPKHSDCLAC	LHFNHSGI	CELHCP	ALVTYNT	DTFES	MPNPEGR	YTFGAS	CVTACP 300
Db	241	AAGCTGPKHSDCLAC	LHFNHSGI	CELHCP	ALVTYNT	DTFES	MPNPEGR	YTFGAS	CVTACP 300
Qy	301	YNYLSTDVGSCTLV	CPHNOE	VTAE	DGTQ	CEKCK	PCARV	CVYGLG	MEHLRE
Db	301	YNYLSTDVGSCTLV	CPHNOE	VTAE	DGTQ	CEKCK	PCARV	CVYGLG	MEHLRE
Qy	361	IOEFAGCKIFGSLA	FLPES	FDGDP	ASNTAP	LOPE	QLOV	FETLEE	ITGYLI
Db	361	IOEFAGCKIFGSLA	FLPES	FDGDP	ASNTAP	LOPE	QLOV	FETLEE	ITGYLI
Qy	421	DLVSFQNLQVIRGR	ILHNGAY	SLTQ	LGIG	ISW	GLRSL	REL	GLGLAL
Db	421	DLVSFQNLQVIRGR	ILHNGAY	SLTQ	LGIG	ISW	GLRSL	REL	GLGLAL
Qy	481	PWDQFRNPHQALL	HTANR	PEDE	CVGEG	LACH	QLCAR	GHCW	GPQPT
Db	481	PWDQFRNPHQALL	HTANR	PEDE	CVGEG	LACH	QLCAR	GHCW	GPQPT
Qy	541	VEECRVLOGLPRE	VYNARH	CLPCH	PECO	PONGS	VTFC	GPEAD	OCVAC
Db	541	VEECRVLOGLPRE	VYNARH	CLPCH	PECO	PONGS	VTFC	GPEAD	OCVAC
Qy	601	PSGVKPDLSYMP	IKWPF	DEEG	ACQPC	INCTH	SCVD	LDKGC	PAEQ
Db	601	PSGVKPDLSYMP	IKWPF	DEEG	ACQPC	INCTH	SCVD	LDKGC	PAEQ
Qy	661	ILLVVLGVVFGIL	IKRRQ	KIRKY	TMRLL	QETEL	VEPL	TPSG	AMPNQ
Db	661	ILLVVLGVVFGIL	IKRRQ	KIRKY	TMRLL	QETEL	VEPL	TPSG	AMPNQ
Qy	721	RKVKVLSGAGF	TVYKGI	WI	PDGEN	VKI	PVAI	KVLR	ENTSP
Db	721	RKVKVLSGAGF	TVYKGI	WI	PDGEN	VKI	PVAI	KVLR	ENTSP
Qy	781	YVSRLLIGICLT	STVQL	VTQ	LM	PGYCL	LDH	VRENR	GR
Db	781	YVSRLLIGICLT	STVQL	VTQ	LM	PGYCL	LDH	VRENR	GR
Qy	841	LVHRDLAARNV	LKSPNHNK	IDT	FGLAR	LLDID	ETEH	ADG	GKVP
Db	841	LVHRDLAARNV	LKSPNHNK	IDT	FGLAR	LLDID	ETEH	ADG	GKVP
Qy	901	HQSDVMSYGVTV	WELMTFG	AKPYD	GIIP	AREIP	DLLE	KGER	LPQ
Db	901	HQSDVMSYGVTV	WELMTFG	AKPYD	GIIP	AREIP	DLLE	KGER	LPQ
Qy	961	IDSECRPRFREL	VSFSESR	MARDP	QRFV	VIQNE	IDLGP	ASPL	DSTFY
Db	961	IDSECRPRFREL	VSFSESR	MARDP	QRFV	VIQNE	IDLGP	ASPL	DSTFY
Qy	1021	BEYLVPQGGFFC	PCPAPG	AGGMV	HRRSS	STRSG	GDLTL	GLP	SEEEA
Db	1021	BEYLVPQGGFFC	PCPAPG	AGGMV	HRRSS	STRSG	GDLTL	GLP	SEEEA
Qy	1081	AGSDVFDGDLG	NGAAKGL	QSLPT	HDPS	PLQRY	SDPTV	PLP	SETD
Db	1081	AGSDVFDGDLG	NGAAKGL	QSLPT	HDPS	PLQRY	SDPTV	PLP	SETD
Qy	1141	NOPDVRPQPP	SPREGPL	PAARPA	AGAT	LERAK	TLSPG	KNGV	VKDV
Db	1141	NOPDVRPQPP	SPREGPL	PAARPA	AGAT	LERAK	TLSPG	KNGV	VKDV
Qy	1201	GGAAPQHPHPP	AFSPAF	NLYYWD	QDPP	ERGAP	STFKGT	PTAEN	PEYIG
Db	1201	GGAAPQHPHPP	AFSPAF	NLYYWD	QDPP	ERGAP	STFKGT	PTAEN	PEYIG


```
Qy 1081 AGSDVFDGDLGMAAKGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
    |||||
Db 1081 AGSDVFDGDLGMAAKGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
    |||||
Qy 1141 NOPDVRPOPPSPRSGPLPAARPAGATLERAKTSLSGKNGVVKDYFAFGGAVENPEYLTPO 1200
    |||||
Db 1141 NOPDVRPOPPSPRSGPLPAARPAGATLERAKTSLSGKNGVVKDYFAFGGAVENPEYLTPO 1200
    |||||
Qy 1201 GGAAPQHPHPPAFSPFDNLYYWDQDPPRGAPPSTFKGTPTAENPEYLGLDVPV 1255
    |||||
Db 1201 GGAAPQHPHPPAFSPFDNLYYWDQDPPRGAPPSTFKGTPTAENPEYLGLDVPV 1255
    |||||
```

Search completed: July 22, 2003, 08:40:17
Job time : 42.9774 secs

THIS PAGE BLANK (USPTO)